Deep Learning for Channel Estimation in RIS-NOMA-assisted THz Communications over Generalized Fading Channels

Tooba Khan, *Student Member, IEEE,* Adnan A. Cheema, *Senior Member, IEEE,* Gökhan Seçinti, *Senior Member, IEEE,* Berk Canberk, *Senior Member, IEEE,* and Trung Q. Duong, *Fellow, IEEE*

***Abstract*—This paper investigates the channel estimation problem in a multiple-input single-output (MISO) reconfigurable intelligent surface (RIS) non-orthogonal multiple access (NOMA)-assisted terahertz (THz) communication system where users experience mobility and varying small-scale fading characterized by the generalized $\alpha - \mu$ distribution. Reliable channel estimation becomes challenging when passive RIS elements and superimposed NOMA signals operate under high attenuation of THz band. We propose a novel deep learning framework, THz RIS-NOMA Channel Estimation (TRiNCE), for cascaded channel estimation. TRiNCE is designed as a GRU-based conditional Wasserstein Generative Adversarial Network with Gradient Penalty (cWGAN-GP). TRiNCE performance is evaluated under various $\alpha - \mu$ fading conditions, the number of RIS elements, NOMA power allocation factors, and the number of BS antennas. The model achieves an R^2 score of 0.85 under Rayleigh fading and up to 0.95 under milder α–μ conditions. It further demonstrates strong robustness, maintaining high estimation accuracy with less than 2% performance variation across different RIS sizes and NOMA power allocation factors, and only $\sim 7\%$ degradation when the number of BS antennas increases from 1 to 5. Results show that TRiNCE not only provides accurate and reliable channel estimation across all tested network configurations but also significantly outperforms convolutional neural network (CNN), long short-term memory (LSTM), and gated recurrent unit (GRU) baseline models while requiring substantially fewer trainable parameters. This establishes TRiNCE as a computationally efficient and effective solution for channel estimation in RIS-NOMA-assisted THz communication systems.**

***Index Terms*—Channel estimation, Deep learning, THz communications, NOMA, RIS, 6G, $\alpha - \mu$ fading, CGAN.**

T. Khan is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1C 5S7, Canada (e-mail: tkhan23@mun.ca).

A. A. Cheema is with School of Engineering, Ulster University, 2-24 York Street, Belfast, BT15 1AP, UK (e-mail: a.cheema@ulster.ac.uk).

G. Seçinti is with Department of Computer Engineering, Istanbul Technical University, 34467 İstanbul, Turkey (e-mail: secinti@itu.edu.tr).

B. Canberk is with is with the School of Engineering and Built Environment, Edinburgh Napier University, EH11 4BN Edinburgh, UK, (e-mail: b.canberk@napier.ac.uk).

T. Q. Duong is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1C 5S7, Canada and also with the School of Engineering and Built Environment, Edinburgh Napier University, EH11 4BN Edinburgh, UK (e-mail: tduong@mun.ca).

This work was supported in part by the Canada Excellence Research Chair (CERC) Program CERC-2022-00109 and in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant Program RGPIN-2025-04941. The work of B. Canberk is partially supported by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) 1515 Frontier R&D Laboratories Support Program for Türk Telekom 6G R&D Lab under project number 5249902.

Corresponding author is Trung Q. Duong.

I. Introduction

Sixth-generation (6G) wireless networks are expected to support new applications and services beyond the scope of 5G, such as holographic communications, immersive 3D multimedia, extended reality (XR), Internet of Everything (IoE), and autonomous systems [1]. These applications will require very high key performance indicators (KPIs), such as peak data rates of up to 1 Tbps, which is fifty times higher than the 20 Gbps supported by 5G, end-to-end latencies below 0.1 ms, and network reliability exceeding 99.99999%. Furthermore, 6G networks will have to support ultradense connectivity with user densities reaching 10^7 devices/km^2 [2]–[6]. This massive connectivity is driven by the paradigm of ubiquitous sensing in smart cities, where emerging data sources, ranging from consumer electronics to large-scale internet of things (IoT) deployments continuously collect and transmit vast amounts of environmental data. The efficient fusion of this multi-modal data is critical for urban resilience but imposes unprecedented demands on the wireless infrastructure, necessitating artificial intelligence (AI)-enhanced solutions to ensure reliable communication. To meet these strict requirements, research has focused on two main approaches: increasing spectral efficiency and exploiting wider bandwidths, typically on the order of several gigahertz (GHz) [7]. However, existing frequency allocations (mmWave bands) offer insufficient bandwidth ($\sim$7 GHz). So, achieving even 0.1 Tbps with this limited spectrum would demand a spectral efficiency of at least 14 bps/Hz, which is unfeasible due to technological and hardware limitations. This limitation necessitates moving beyond mmWave toward higher frequency bands [8].

The largely untapped terahertz (THz) band spanning 0.1–10 THz offers wide bandwidth availability, providing the capacity for extremely high data rates (exceeding 100 Gbps), to support the massive traffic expected in 6G networks [1], [9], [10]. Moreover, the diminishing wavelength in the THz band helps in the development of high gain antennas, enabling massive spatial multiplexing and improved communication efficiency [11]. The ultra high operating frequency also reduces the transmission latency, which is critical for delay sensitive applications [9].

However these shorter wavelengths also present some significant challenges. THz waves experience higher Free Space Path Loss (FSPL) and increased molecular absorption, especially due to atmospheric gases and weather-related attenua-

tion, which can severely limit their effective range [9], [12]. Reflection and scattering losses are also more pronounced in THz frequencies because the short wavelengths easily reflect from rough surfaces and human blockages [13], [14]. To address these propagation challenges, emerging technologies like reconfigurable intelligent surfaces (RISs) [15] and non-orthogonal multiple access (NOMA) [16] have the potential to enhance coverage, spectral efficiency, and energy efficiency [17] while leveraging the unique advantages of the THz band.

RIS is a two-dimensional array of low cost and passive elements that offers programmable control over the wireless environment [18]. By controlling the phase shift of these elements, RIS creates an adaptive signal steering environment which can boost the signal coverage and minimize the interference [19]–[22]. In order to mitigate signal blockage in THz communication, RIS can create virtual line-of-sight (LoS) links between base stations and users where direct paths are blocked, thereby improving data rates and reliability [23], [24]. The passive nature of these RIS elements results in low power consumption, offering a sustainable solution for enhancing system capacity [25]. Furthermore, RIS can be easily installed on existing infrastructure, such as walls, buildings, or billboards, making it a practical and cost-effective deployment strategy for dense urban environments [24].

Another important enabler for 6G is NOMA, which helps to increase spectral efficiency by enabling multiple users to share the same time-frequency resources by using power-domain multiplexing. While traditional orthogonal multiple access (OMA) schemes need to reserve separate resource blocks for each user, NOMA uses intelligent power allocation and successive interference cancellation (SIC) to enable massive connectivity, one of the most important requirements for ultra-dense networks of 6G [26]. This approach helps to improve system capacity and the quality of service (QoS) by dynamically prioritizing cell edge users by optimized power distribution [27]. Furthermore, NOMA's inherent flexibility enables efficient support for diverse 6G applications with varying QoS requirements, from bandwidth-intensive extended reality to low-power massive IoT deployments [2], [28].

Recent research shows that the integration of RIS with NOMA in THz communications has emerged as a promising strategy to enhance wireless network performance. By combining the strengths of both technologies, with the high bandwidth availability at THz frequencies, we can address the critical demands of next-generation wireless systems, including spectrum scarcity, high data rate requirements, energy efficiency, and severe path loss inherent to THz frequencies [29]. A RIS-NOMA-assisted user pairing scheme for THz communication (RTHz-NOMA), proposed in [30], shows that RTHz-NOMA significantly outperforms OMA in terms of sum-rate performance, even under molecular absorption conditions. Furthermore, the proposed user-pairing approach leads to a substantial enhancement in the bit error rate (BER) when compared to conventional THz NOMA and OMA systems that do not incorporate RIS. RIS assisted NOMA in THz networks significantly improves aggregated throughput and spectral efficiency compared to conventional schemes, especially with joint optimization of power allocation and RIS beamforming [23], [31], [32]. The use of advanced RIS configurations like STAR-RIS further improves energy efficiency in multiple-input and multiple-output (MIMO)-NOMA systems when compared to OMA and random phase shift-based schemes [32]–[35]. It is worth noting that the performance gains in spectral and energy efficiency offered by the RIS-NOMA-THz framework come at the cost of increased system complexity, particularly regarding channel estimation and SIC. However, these technologies are critical for overcoming the severe blockage and connectivity challenges of THz-based 6G networks. Consequently, the development of low-complexity algorithms for channel estimation is essential to mitigate this overhead, ensuring that these system trade-offs remain favorable for next-generation networks.

Due to the short wavelengths at THz band, signals are highly sensitive to scattering, reflection, and rapid constructive and destructive interference, which results in significant small-scale fading. This makes accurate fading models essential for robust system design [13], [14]. Given the complex propagation environment of the THz band, accurate modeling of small-scale fading becomes difficult. Most of the research in THz channel modelling has been conducted with a LoS channel considering deterministic large-scale fading effects only [36]. In [36], a simplified molecular absorption loss model is proposed for 100-450 GHz bands. Recent research is being done to investigate the small-scale fading effects of THz bands [37]. Traditional fading models like Rayleigh and Rician are often unsuitable for THz frequencies, whereas the more general $\alpha-\mu$ distribution [38] has been shown to fit empirical measurements more accurately, confirmed by the Kolmogorov-Smirnov (KS) test [37], [39]. The $\alpha-\mu$ distribution is a flexible fading model that includes the Gamma, exponential, Weibull, Nakagami-m, Rayleigh, and one-sided Gaussian distributions as special cases. While traditional models like Rayleigh and Rician are simpler and offer mathematical tractability, they are unable to capture the complex propagation characteristics of THz links [39]. In contrast $\alpha-\mu$ fading model provides an excellent fit for small-scale fading in both LoS and non-line-of-sight (NLoS) THz wireless links [37], [39]. Moreover, it also maintains mathematical tractability, which allows for the derivation of closed-form expressions for key performance metrics such as outage probability, BER, and ergodic capacity, which is essential for evaluating and optimizing THz communication systems [40]–[43]. Numerous studies have validated these analytical results through simulations, confirming the $\alpha-\mu$ distribution's accuracy and effectiveness in THz channel modeling [40], [41]. The main contributions of this paper are summarized as follows:

- We propose and investigate a novel, integrated system model for a downlink MISO RIS-NOMA-assisted THz communication system. The model incorporates multiple transmit antennas at the BS and multiple single-antenna users, enabling a more practical and scalable multi-user setting compared to existing SISO-based works. The channel is modeled under the generalized α–μ fading distribution, enabling robust evaluation across diverse small-scale fading conditions.

- We develop a comprehensive time-varying MISO RIS-NOMA-THz dataset generation pipeline that produces complex in-phase and quadrature (I/Q) components of received signals and the corresponding complex I/Q cascaded channels. The dataset incorporates multiple users, $\alpha - \mu$ fading, FSPL, molecular absorption effects, and user mobility.
- We propose TRiNCE, a lightweight GRU-based conditional WGAN-GP architecture specifically designed for RIS-NOMA-assisted THz channel estimation. TRiNCE efficiently learns the nonlinear mapping from received complex I/Q NOMA signals to the complex-valued cascaded channels, capturing both real and imaginary components while maintaining low computational complexity to mitigate the trade-off between system performance and feasibility.
- TRiNCE is extensively evaluated across variations in fading severity controlled by $\alpha - \mu$, number of RIS elements, NOMA power allocation factors and number of transmit antennas.
- We demonstrate that the proposed model outperforms convolutional neural network (CNN), long short-term memory (LSTM), gated recurrent unit (GRU) and CNN-LSTM models. It achieves highest R^2 score and lowest normalized mean squared error (NMSE) with significantly fewer trainable parameters than competing models.

The rest of the paper is structured as follows. Section II surveys the related works concerning THz propagation channel modeling, RIS-NOMA integration in THz communications, and existing channel estimation techniques in RIS-NOMA and RIS-assisted THz systems. Section III presents the downlink RIS-NOMA-assisted THz communication system model and formulates the channel estimation problem. The proposed deep learning (DL) framework is introduced in Section IV. Section V presents the performance analysis of the proposed model under various channel conditions, numbers of RIS elements, NOMA power allocation factors and number of transmit antennas. Finally, Section VI concludes the paper and outlines potential directions for future work.

TABLE I: Summary of channel estimation techniques for RIS-NOMA-THz systems

Ref.	RIS	NOMA	THz	Small-scale fading
[44]	✓	✓	–	Rician
[45], [46]	✓	✓	–	Rayleigh
[47], [48]	✓	✓	–	Rayleigh
[49]	✓	✓	–	Rayleigh
[50], [51]	–	–	✓	Geometric
[52], [53]	✓	–	✓	Geometric
[54]–[56]	✓	–	✓	Geometric
This work	✓	✓	✓	$\alpha - \mu$

II. Related Works

A. Channel Modelling in THz band

Channel modeling is the process of mathematically expressing the channel characteristics and serves as the basis for system design, optimization, and performance analysis. There are two types of channel models: deterministic models and stochastic models [57]. Deterministic models are based on precise environmental data, e.g., the building's architecture, and physical laws governed by electromagnetic wave theory to determine how the signal will interact with and travel through the environment. Key deterministic effects include FSPL and molecular absorption caused by atmospheric gases. Stochastic channel models use random variables and statistical distributions, e.g., Rayleigh, Rician, $\alpha - \mu$, to capture the unpredictable complex channel characteristics. These models are based on the statistical properties from real measurements.

Several deterministic and statistical models have been proposed to capture the characteristics of THz channels. The work in [58]–[60] explores the effects and models for FSPL and molecular absorption in THz channels. A simplified model to capture the deterministic fading effects in the 200 to 400 GHz bands is presented in [61]. This model is adapted for the RIS-aided communications in [62]. An analytical path loss model for the THz band, dependent on frequency, propagation distance, and altitude, is proposed in [63]. According to the comparative analysis of three different path loss models, two statistical and one deterministic, done in [60], the deterministic path loss model showed much higher path loss than the statistical models because it was able to capture the environment more accurately. In [64], the authors present a survey of three main aspects of the THz channel model: experimental channel sounding, deterministic ray tracing simulations, and statistical stochastic channel modeling.

Geometry-based stochastic models (GBSMs) inherently capture both large-scale and small-scale fading by combining deterministic cluster geometry with stochastic multipath effects. A 3D GBSM for dynamic MIMO channels in the sub-THz band is presented in [65]. The model uses a receiver centered spherical geometry with local clusters of scatterers and individual reflectors. Simulations show that speed, motion duration, antenna spacing, and carrier frequency are important factors to be considered for designing future sub-THz uncrewed aerial vehicle (UAV)-Air-to-Air(A2A) MIMO systems. Another 3D GBSM model is proposed in [66] that uses a double spherical structure to represent the wireless channel. By incorporating the effect of nano-material antenna arrays and the high losses of THz bands, the model captures key THz ultra massive MIMO characteristics. The authors concluded that the proposed model and its statistical properties are useful for designing THz systems for 6G and beyond. Geometric channel models have also been used to characterize the cascaded channels in RIS-aided THz communications in [52]–[56].

Small-scale fading modeling in the THz band is a relatively new area of research, with most of the literature being quite recent. Experiments in [67] show that the fading envelope follows the Nakagami-*m* distribution under both LoS and NLoS conditions. Rician fading models have been selected for THz channel modelling in [11], [29]. In [11], the authors assumed that THz communications are dominated by LoS propagation paths, making Rician fading more appropriate to model the effects of small-scale fading. The authors in [29] consider a channel model that incorporates both LoS compo-

nent and NLoS multipath components. The LoS component of the channel is modeled using Rician fading, and the individual NLoS paths are modeled using Rayleigh distributed amplitude variations combined with distinct time delays. However, measurement campaigns show that traditional fading models such as Rayleigh, Rician, or Nakagami-*m* are generally insufficient to capture the characteristics of THz signals under both LoS and NLoS conditions [39]. To address this, more flexible models have been considered. The fluctuating two-ray (FTR) model [68] has been applied in RIS-THz systems and THz backhaul networks to derive analytical expressions for outage probability and ergodic capacity [62], [69], providing better flexibility in modeling random fluctuations of the signals. However, the high costs of numerical evaluations associated with the FTR model make it computationally inefficient [70]. Among generalized fading models, the α–μ distribution has become a popular choice because of its ability to fit THz channel measurements more accurately than classical models [39]. Inspired by the work done in [39], the authors in [71] select α–μ fading model to develop a comprehensive analytical framework for RIS-assisted THz systems. They propose a novel closed-form approximation that eliminates the need for computationally complex multivariate Fox's H functions while maintaining mathematical tractability. Their results support that the α–μ modeling is essential for accurate performance assessment of THz systems. In [40], the performance of a RIS-assisted THz system is analyzed by deriving exact closed-form expressions for the probability density function (PDF) and cumulative distribution function (CDF) of the small-scale fading channel, considering the combined effects of α–μ fading and zero-boresight pointing errors. In [72], the α–μ distribution is used to capture the statistical variations of THz signals in indoor STAR-RIS-assisted NOMA systems, while for outdoor settings, mixture-based models (Gamma and Gaussian mixture) are adopted. Based on these fading models, outage probability and ergodic capacity expressions are obtained for STAR-RIS-assisted THz systems and confirmed the suitability of the α–μ distribution as a generalized framework for small-scale fading in THz communications.

B. Existing Techniques for Channel Estimation

For NOMA systems, accurate channel state information (CSI) is essential at the receiver to ensure reliable detection and decoding of each user's data, and in dynamic environments with fading or mobility, estimation errors can significantly degrade performance [44], [73]. Channel estimation in RIS-NOMA systems is challenging due to the passive nature of RIS elements and the presence of superimposed signals inherent to NOMA. In [45], a CNN-LSTM-based DL model for robust channel estimation in RIS-NOMA systems is proposed. It combines CNN for spatial feature extraction and LSTM for temporal feature learning. This method outperforms traditional models, resulting in improved estimation accuracy across multiple performance metrics. The proposed model utilizes both the magnitude and phase of the received signal as inputs and estimates the channel magnitude. The study in [46] extends this approach using Quantum-based LSTM(QLSTM) models in to highlight the potential of quantum machine learning(QML) in channel estimation. A hybrid architecture combining a 1D-CNN with Bidirectional LSTM (BiLSTM) and GRU was proposed for estimating NOMA channel magnitudes in [74]. Linear minimum mean square error (LMMSE) estimation was used to estimate the complex channel gains in RIS-NOMA setups [44]. They found that increasing the pilot sequence length improves estimation accuracy even with hardware impairments. Techniques like alternating least squares (ALS) and vector approximate message passing can estimate the cascaded channels iteratively, achieving performance close to the theoretical Cramér-Rao Bound [49].

THz channels have high propagation losses and strong frequency selectivity making accurate channel estimation even more difficult [75], [76]. THz channel estimation is an emerging research domain, and the majority of existing literature has focused on the estimation of geometric channel models. A CNN based THz channel estimator that estimates real-valued channel factors was proposed in [50]. These factors include path loss, time of arrival, phase change on arrival and azimuth and elevation angles of arrival and departure. Using these estimated factors with the geometric model equation, the channel can be determined completely. A tensor based channel estimator was used in [54] to estimate the geometric structure of the RIS-THz channel. This method provides accurate parameter estimates and maintains computational efficiency [54]. Recent studies have explored the sparse structure of the THz channel in the angular domain to improve channel estimation accuracy. A Two-stage estimator leverages the sparsity in the angular domain of the cascaded channels to separate the BS-RIS and RIS-user channels, using atomic norm minimization and low-complexity algorithms like alternating direction method of multipliers (ADMM) to achieve superior estimation performance compared to traditional approaches [52]. The authors in [77] proposed a sparsifying matrix learning-aided Bayesian approach for THz ultra-massive (UM)-MIMO systems. This method utilizes dictionary learning to capture the coexistence of near- and far-field paths, effectively handling the structured sparsity of the channel. Compressed sensing (CS) based methods have also been proposed for channel estimation to reduce pilot overhead and computational complexity by exploiting the dual sparsity of THz MIMO channels in both angular and delay domains [56], [78].In [55], the authors exploit the inherent sparsity of the THz geometric channel and formulate channel estimation as a sparse recovery problem. They replace traditional CS algorithms with a two-stage neural network. In the first stage, two linear perceptrons separately process the real and imaginary parts of the received pilot signals to mimic the linear measurement process. Second stage uses a multilayer perceptron (MLP) to estimate the sparse channel gains of the propagation paths. In [79] a Markov process based THz channel is estimated using a residual network, Estimation ResNet (E-ResNet), and the model is evaluated for a single user. A deep learning model for channel estimation in hybrid-field THz UM-MIMO systems was proposed in [51], using geometric models for near and far field channels. The network has a BiLSTM layer, a GRU layer, and an output dense layer with a single neuron. For a single antenna, the received signal

has N complex pilot samples, which are split into real and imaginary parts, giving $2N$ real values. Each value is treated as a separate timestep by the proposed model, which processes the sequence to predict a real-valued channel representation. This procedure is repeated for all N antennas to obtain the full set of channel estimates.

In RIS-NOMA-assisted THz systems, accurate estimation of the cascaded channel presents a critical challenge. Reliable channel knowledge is essential for effective SIC to decode superimposed user signals. However, to avoid the high overhead of conventional pilot-based methods, the receiver must infer the cascaded channel directly from the received NOMA signals. This task is complicated by THz-specific impairments such as severe path loss, molecular absorption attenuation, and rapidly fluctuating small-scale fading. Additionally, the cascaded channel consists of multiple antenna paths, each with I and Q components, which increases the dimensionality and requires the simultaneous prediction of multiple complex channel components.

A critical analysis of existing literature, summarized in Table I, reveals a significant gap in this integrated context. While channel estimation for RIS-NOMA systems has been extensively explored in works such as [44], [45], and [46], these studies are limited to conventional fading environments. Specifically, they rely on standard Rayleigh or Rician fading models, which fail to capture the severe path loss and unique propagation characteristics inherent to the THz band. Recent advancements in THz channel estimation include DL based geometric model estimators in [50] and [51], which focus on THz channels without RIS. For RIS-assisted THz systems, tensor-based approaches in [54] and DL based geometric model estimators in [55] are also proposed. In these existing approaches, the network either predicts only the channel magnitude or estimates real-valued geometric channel factors, which then need to be combined using geometric equations to reconstruct the full complex channel. Moreover, existing THz channel estimation works do not account for the complexity of superimposed signals found in NOMA and rely heavily on deterministic geometric channel models rather than generalized statistical distributions. In contrast, the proposed method directly predicts both the real and imaginary components of the complex channel, enabling end-to-end complex channel estimation. Consequently, no prior work has addressed channel estimation within a fully integrated RIS-NOMA-THz framework, nor has the performance of such a system been evaluated under the generalized $\alpha-\mu$ fading model, which provides a more versatile representation of the scattering and non-linearity observed in high-frequency communications.

III. System Model and Problem Formulation

We consider a downlink communication system that uses RIS-NOMA in the THz band over α–μ fading channels, as illustrated in Fig. 1. The system architecture has a base station (BS) with N_T antennas, a RIS with M reflecting elements, and J single-antenna user equipments (UEs), enabling a multi-user MISO architecture. In this scenario, it is assumed that both the BS and the RIS are fixed, while only the UEs are slowly moving away from the BS. Additionally, each RIS element is presumed to be sufficiently distant to guarantee signal uncorrelation and minimize mutual signal interference. Furthermore, since the direct channel between the BS and the UE is assumed to be blocked by obstacles, the signal is received at each UE through a cascaded channel consisting of the BS to RIS channel and the RIS to UE channel. Throughout this paper, we assume that the wireless channels are quasi-static within a coherence block of duration T_c. For THz communications, the high carrier frequency and slow UE movement result in a coherence time on the order of a few milliseconds. The complete cascaded channel matrix from the BS to all UEs via the RIS at time t is denoted by $\mathbf{G}(t) \in \mathbb{C}^{J\times N_T}$, whose j^{th} row corresponds to the per-user cascaded channel $\mathbf{g}_j(t) \in \mathbb{C}^{1\times N_T}$ [80], [81], and is calculated as

$$\mathbf{g}_j(t) = \rho_{LS}(t)\mathbf{k}_j^H(t)\mathbf{\Phi}_j(t)\mathbf{H}(t), \tag{1}$$

here, the diagonal coefficient matrix $\mathbf{\Phi}_j(t)$ of size $M \times M$ is defined as $diag\{\phi\}$ with $\phi = \Lambda_1 e^{j\theta_1}, \Lambda_2 e^{j\theta_2}, \cdots, \Lambda_M e^{j\theta_M}$ where Λ_m represents the amplitude reflection coefficient and $\theta_m \in [0, 2\pi]$ represents the phase shift for the m^{th} RIS element. The large-scale fading component $\rho_{LS}(t)$ captures both the propagation loss and molecular absorption, and is calculated as [62]

$$\rho_{LS}(t) = \frac{c^2\sqrt{\mathcal{A}_{\mathcal{G}t}\mathcal{A}_{\mathcal{G}r}}}{(4\pi f)^2(d_1 d_{2,j})} e^{-\frac{1}{2}\varrho(f)(d_1+d_{2,j})}, \tag{2}$$

where c is the speed of light, $\mathcal{A}_{\mathcal{G}t}$ and $\mathcal{A}_{\mathcal{G}r}$ are the antenna gains at the transmitter and receiver, respectively, f is the operating frequency in GHz, d_1 is the distance in meters from the BS to the RIS, $d_{2,j}$ is the distance in meters from the RIS to UE_j, and $\varrho(f)$ denotes the molecular absorption coefficient. The molecular absorption coefficient varies with the operating frequency [71]. For the 275–400 GHz range, it is determined according to the absorption loss model in [61].

The channel matrix from the BS to the RIS is represented by $\mathbf{H}(t) \in \mathbb{C}^{M\times N_T}$. As both the BS and RIS are fixed, this link is assumed to be time-invariant. The channel matrix from the RIS to UE_j is denoted as $\mathbf{k}_j(t) \in \mathbb{C}^{M\times 1}$, and $\mathbf{k}_j^H(t)$ is its Hermitian transpose. The small-scale fading components of both these channel matrices are modeled using the α–μ distribution, which effectively captures the wireless channel's nonlinear and clustered multipath characteristics. The PDF of the α–μ fading envelope is given by

$$f_{\rho_{SS}}(x) = \frac{\alpha\beta^{\alpha\mu}}{\Omega^{\alpha\mu}\Gamma(\mu)} x^{\alpha\mu-1} e^{-(\beta\frac{x}{\Omega})^{\alpha}}, \tag{3}$$

here $\alpha > 0$, known as the power parameter, controls the nonlinearity of the propagation medium and shapes the fading distribution. The parameter $\mu > 0$ denotes the number of multipath clusters and characterizes the clustering intensity of scattered waves [38]. Both α and μ parameters govern the severity of fading. An increase in either parameter, while the other is kept constant, corresponds to less severe fading, while a decrease in either parameter results in more severe fading [43]. The parameter Ω denotes the expected value of the signal envelope, and $\Gamma(\cdot)$ is the Gamma function. The normalization factor β is given by $\beta = \Gamma(\mu + 1/\alpha)/\Gamma(\mu)$ [38].

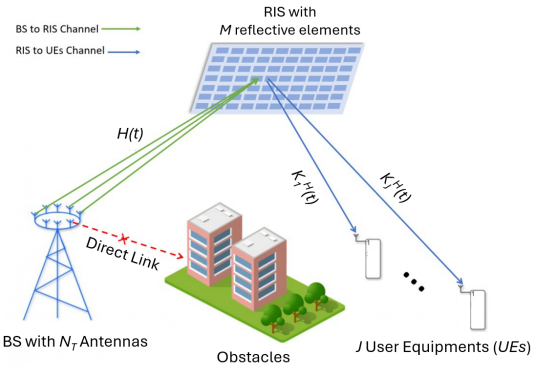


Fig. 1: System model for a down-link RIS-NOMA-assisted THz communication system

Singular value decomposition (SVD) based optimal beamforming [82] is used at the BS to transmit the superimposed NOMA signal toward the RIS. The BS is assumed to have perfect knowledge of the cascaded channel $BS-RIS-UE_j$. This CSI is used to compute the beamforming vector $\mathbf{w}(t) \in \mathbb{C}^{N_T \times 1}$, by solving the optimization problem:

$$\max_{\mathbf{w}(t)} |\mathbf{H}(t)\mathbf{w}(t)|^2 \quad \text{s.t.} \quad |\mathbf{w}(t)|^2 = 1. \tag{4}$$

The solution is derived from the SVD of $\mathbf{H}(t) = \mathbf{U}(t)\mathbf{\Sigma}(t)\mathbf{V}^H(t)$, where the optimal beamformer aligns with the strongest spatial mode of the channel,

$$\mathbf{w}(t) = \mathbf{v}_1(t), \tag{5}$$

where $\mathbf{v}_1(t)$ is first column of $\mathbf{V}(t)$, the principal right-singular vector of $\mathbf{H}(t)$, corresponding to its largest singular value σ_1. Here $\mathbf{w}(t)$ is a unit norm vector satisfying the transmit power constraint, and this design maximizes the received power at the RIS since $\|\mathbf{H}(t)\mathbf{v_1}(t)\| = \sigma_1^2$, enabling efficient reflection to all UEs via the RIS.

According to the principles of NOMA, the superimposed signal, $x(t)$, combines the information for all J users and is given by [83]

$$x(t) = \sum_{j=1}^{J} \sqrt{P_s \omega_j} s_j(t), \tag{6}$$

where P_s is the total transmitted signal power, ω_j is the power allocation factor of jth UE, and s_j is the signal transmitted for UE_j. We consider that the users are ordered based on their channel gains such that near user, UE_1, has the strongest channel and is assigned the smallest power allocation factor, while far user, UE_J, has the weakest channel and is assigned the largest power allocation factor. The power allocation coefficients are such that $\sum_{j=1}^{J} \omega_j = 1$. The final signals transmitted from the BS antennas are $\mathbf{x}(t) = x(t)\mathbf{w}(t)$.

At the receiver side, at every time instant, the signal received at each user is expressed as

$$y_j(t) = \mathbf{g}_j(t)\gamma \frac{\mathbf{x}(t)}{\sqrt{P_s}} + n_j(t), \tag{7}$$

where $\gamma = P_s/N_0$ is the transmit SNR, with N_0 denoting the noise power. The noise power N_0 is computed assuming a transmission bandwidth of 1 GHz and a receiver noise figure N_F of 5 dB. The receiver noise $n_j(t)$ at each UE is modeled as a zero-mean complex Gaussian random variable with variance N_0 [84].

The dataset generation process aims to simulate a two user downlink RIS-NOMA-THz system under time-varying channel conditions due to UE mobility. This sequential dataset is generated for T time steps.The channel matrices $\mathbf{H}$, $\mathbf{k}_1(t)$ and $\mathbf{k}_2(t)$ are generated using the α–μ fading distribution with parameters α, μ, Ω. UE mobility at each time step is simulated by incrementing the RIS-UE distance with v. The transmitted NOMA signal combines Gaussian-distributed symbols $s_1(t), s_2(t) \sim \mathcal{CN}(\mu_{sj}, \sigma_{sj}^2)$ with power allocation factors ω_1, ω_2. Over T iterations, the received signals $\mathbf{Y}(t)$ and channel matrices $\mathbf{G}(t)$ are collected to form the dataset. To create diverse training and testing environments the system parameter including the number of BS antennas N_T, the number of RIS elements M, the fading parameters α and μ, and the NOMA power allocation factors ω_1 and ω_2 are varied across different dataset configurations. The complete workflow is mentioned in Algorithm 1.

IV. Proposed Model

In conventional DL based channel estimation, models such as CNNs or RNNs are trained to map the received signal to a single deterministic channel realization. However, this may lead to poor generalization in highly dynamic scenarios, such as those encountered in RIS-NOMA-assisted THz communication systems, where channel characteristics are affected by severe fading, cascaded propagation paths, and noise.

To address these limitations, we propose TRiNCE (THz RIS–NOMA Channel Estimation), a GRU-based conditional Wasserstein GAN for channel estimation in RIS–NOMA-THz communications. Unlike traditional DL models that directly predict the channel estimates, TRiNCE learns the conditional distribution of the cascaded BS–RIS–UE channel given the received signal at the UE by training the generator with a combination of L2 and Wasserstein losses. Initially, the L2 loss dominates to guide the generator towards accurate channel estimates and later Wasserstein loss takes over to capture the full stochastic behavior of the cascaded channel conditioned on the received signal, effectively modeling $p(\mathbf{G} \mid \mathbf{Y})$.

This section describes the proposed TRiNCE framework for RIS–NOMA assisted THz communications, simulation setup and the data preprocessing steps.

A. Proposed TRiNCE Model

Conventional Generative Adversarial Networks (GANs) [85], which rely solely on noise as input to the generator, are effective at generating realistic-looking data. However, for wireless channel estimation, generating realistic-looking channels alone is insufficient. The generated data must also capture the temporal dependencies and dynamic behavior of the wireless channels. Therefore, we use a conditional GAN (CGAN) [86], which extends the conventional GAN by providing conditional inputs to both the generator and discriminator so that the generation process is guided by the received signal,

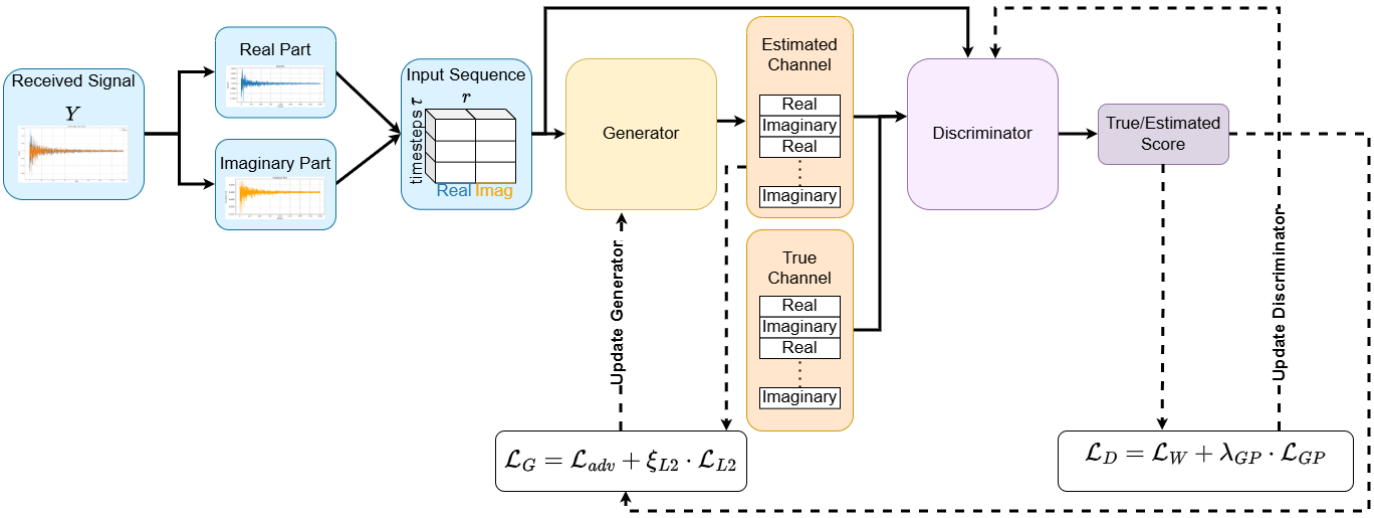


Fig. 2: Proposed Estimation model.

resulting in channel realizations consistent with the input. Despite this, CGANs still share some limitations of conventional GANs, including vanishing gradients, mode collapse, and unstable training behavior [87]. To address this, a Conditional Wasserstein GAN with Gradient Penalty (cWGAN-GP) [88] based architecture is proposed for channel estimation. In this conditional adversarial learning setup, the generator aims to produce estimated channel coefficients that accurately capture the true underlying channel, while the discriminator learns to distinguish between true and generated channel samples. The model integrates the Wasserstein-1 distance as the adversarial objective function, providing smoother and more meaningful gradients compared to the conventional GANs. Additionally, the gradient penalty term is used to enforce the Lipschitz constraint, ensuring stable convergence during training and mitigating the problems of mode collapse and gradient explosion [87].

The proposed TRiNCE framework is illustrated in Fig. 2. During training phase, the discriminator is updated n_d times for every n_g generator update to stabilize adversarial learning and prevent the generator from overpowering the discriminator too early in the training process. This framework enables the generator to slowly learn to produce channel estimates that closely resemble the true channel distribution, while the discriminator continues to learn to distinguish between true and generated samples. Once the generator is trained, it is used to directly estimate the channel coefficients from the received signal sequences without requiring any further adversarial updates.

The generator, shown in Fig. 3, takes as input a sequence of past τ received signal samples $r = [\Re\{Y[t-\tau+1:t]\}, \Im\{Y[t-\tau+1:t]\}] \in \mathbb{R}^{\tau \times 2}$, and produces the corresponding channel estimates, $\hat{g}$, for target channel coefficients g. Each input sample is represented by two values, corresponding to the real and imaginary parts of the received signal. Two stacked GRU layers are used to capture the temporal dependencies present in the input sequence. GRUs use update and reset gates to control the flow of information, allowing

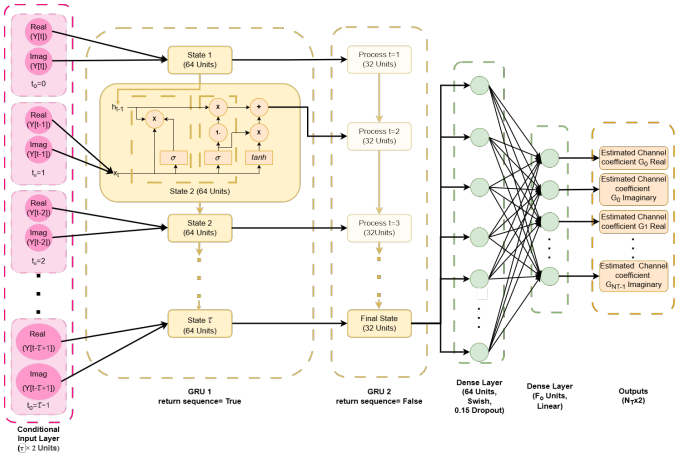


Fig. 3: Generator architecture.

them to efficiently retain useful information while forgetting non-useful information [89]. GRU is computationally simpler and faster to train as compared to LSTM [90].

The temporal features extracted by the GRU layers are passed through a fully connected dense layer with 64 units and the swish activation function, followed by a dropout layer to prevent overfitting. The dense layer serves as a feature transformation stage, combining the learned temporal features into a more discriminative form before generating the channel estimate. The swish activation function [91], defined as $f(x) = x \cdot sigmoid(x)$, is designed to improve over the Rectified Linear Unit (ReLU) activation function. Unlike ReLU, it introduces a smooth and non-monotonic behavior and self-gating properties that help the network maintain small negative activations instead of abruptly zeroing the neurons, leading to improved convergence and performance in deep networks [91]. Finally, a dense output layer with a linear activation function generates F_o estimated channel coefficients. Here, $F_o = N_T \times 2$, representing the real and imaginary components of the cascaded channel for each of the transmit antennas.

Algorithm 1: RIS-NOMA Dataset Generation

Input: $N_t, M, J, T, v, P_s, f, \alpha, \mu, \Omega, \omega_1, \varrho(f), d_1, d_{1,1}, d_{1,2}, \mathcal{A}_{\mathcal{G}_t}$ and $\mathcal{A}_{\mathcal{G}_r}$

Output: Dataset $\mathcal{D} = \left\{ \left(Y_{t,j}, \{G_{t,j,n_t}\}_{n_t=1}^{N_t} \right) \right\}_{j=1,t=1}^{J,T}$

Initialize the simulation parameters according to Table II

Initialize $\mathbf{Y} \in \mathbb{C}^{T \times J}$, $\mathbf{G} \in \mathbb{C}^{T \times J \times N_T}$ as empty matrices

Generate time-invariant channel matrix $\mathbf{H} \in \mathbb{C}^{M \times N_T}$ using α, μ, Ω

Compute the beamforming vector $\mathbf{w} \in \mathbb{C}^{N_T \times 1}$ using eq. (4) and eq. (5)

Compute the noise power N_0 using bandwidth B and noise figure N_F

for $t = 1$ *to* T **do**

  Generate $\mathbf{k_1} \in \mathbb{C}^{M \times 1}$ using α, μ, Ω

  Generate $\mathbf{k_2} \in \mathbb{C}^{M \times 1}$ using α, μ, Ω

  Compute $\mathbf{k_1^H} \in \mathbb{C}^{1 \times M}$ and $\mathbf{k_2^H} \in \mathbb{C}^{1 \times M}$ by taking the Hermitian transpose of $\mathbf{k_1}$ and $\mathbf{k_2}$

  Generate s_1 and s_2 as complex Gaussian signals, $\mathcal{CN}(1, 0.1)$

  Compute the superimposed signal x using eq. (6)

  Compute the precoded signal $\mathbf{x} = x \cdot \mathbf{w}$

  Initialize $\mathbf{\Phi}_j \in \mathbb{C}^{M \times M}$ as an empty matrix

  for $j = 1$ *to* J **do**

    Generate n that follows $\mathcal{CN}(0, N_0)$

    Generate phase vector $\boldsymbol{\theta}[t] \sim \mathcal{U}(0, 2\pi)^M$

    Construct $\boldsymbol{\phi}[t] = [\Lambda_1 e^{j\theta_1[t]}, \Lambda_2 e^{j\theta_2[t]}, \ldots, \Lambda_M e^{j\theta_M[t]}]$

    Using $\boldsymbol{\phi}[t]$ compute $\mathbf{\Phi}_j[t] = \text{diag}\{\boldsymbol{\phi}[t]\}$

    Update distance: $d_{2,j}[t] = d_{2,j}[t-1] + v$

    Compute path loss $\rho_{LS}[t]$ from eq. (2)

    for $n_t = 1$ *to* N_T **do**

      Compute $\mathbf{G}[t, j, n_t]$ from eq. (1)

    end

    Compute $\mathbf{Y}[t, j] = \sum_{n_t=1}^{N_T} (\mathbf{G}[t, j, n_t] \cdot \gamma \cdot \frac{\mathbf{x}[n_t, t]}{\sqrt{P_s}}) + \mathbf{n}[t, j]$

  end

end

Dataset $\mathcal{D} = \left\{ \left(\mathbf{Y}_{t,j}, \{\mathbf{G}_{t,j,n_t}\}_{n_t=1}^{N_t} \right) \right\}_{j=1,t=1}^{J,T}$

The generator framework is trained using a joint loss function that balances adversarial training with reconstruction loss. The total generator loss $\mathcal{L}_G$ is formulated as

$$\mathcal{L}_G = \mathcal{L}_{adv} + \xi_{L2} \cdot \mathcal{L}_{L2}, \tag{8}$$

where $\mathcal{L}_{adv}$ is the adversarial loss and $\mathcal{L}_{L2}$ is the reconstruction loss. $\mathcal{L}_{adv}$ encourages the generator to produce channel estimates that accurately capture the stochastic behavior of the cascaded channel, aligning the generated distribution with the true channel distribution, while $\mathcal{L}_{L2}$ ensures that the predicted channel has minimum prediction error.

$\mathcal{L}_{L2}$ is weighed by a dynamic parameter ξ_{L_2}. The value of ξ_{L_2} is decayed over epochs to balance the adversarial training and reconstruction accuracy. Initially, the generator produces estimates that are far from the real data distribution. So a higher value of ξ_{L_2} forces the generator to learn the true mapping between the past sequence of received signals r from the distribution of received signal, p_Y, and target channel coefficients g from the distribution of cascaded channel p_G. This gives the generator a better starting point and prevents overfitting while maintaining robustness against noise in received signals [92]. Over the epochs as the generator and discriminator are trained, $\mathcal{L}_{adv}$ becomes more meaningful and starts guiding the generator and are expressed as

$$\mathcal{L}_{adv} = -\mathbb{E}_{r \sim p_Y}[D(r, \hat{g})]. \tag{9}$$

$$\mathcal{L}_{L2} = \mathbb{E}_{(r,g) \sim p_{Y,G}}[\|g - \hat{g}\|_2^2]. \tag{10}$$

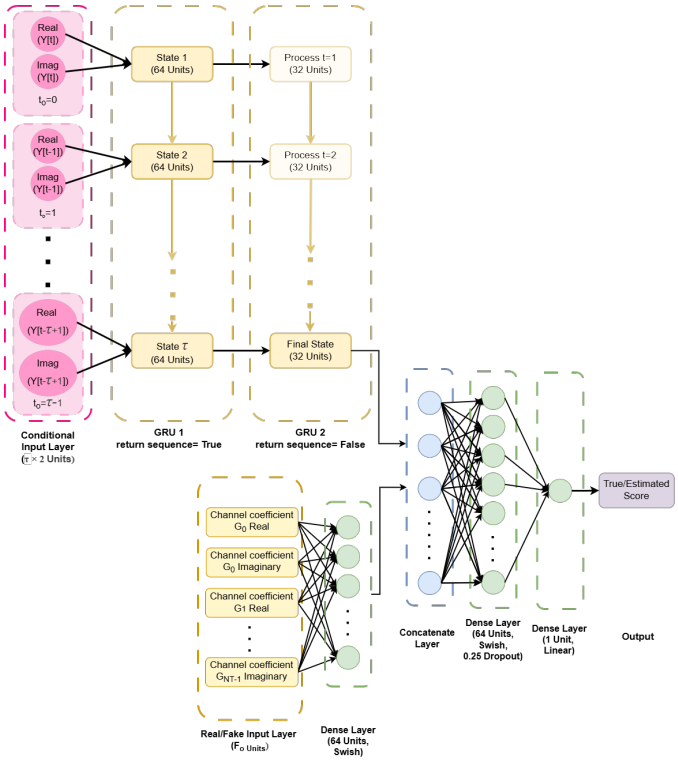


Fig. 4: Discriminator architecture.

The discriminator architecture, shown in Fig. 4 is designed to distinguish between true channel samples and generator-produced estimates by conditioning on both the received signals and channel coefficients. The received signal sequence is processed through two GRU layers to extract temporal features, while the channel input is passed through a dense layer with swish-activated units. Next, the extracted features from both are combined to form a joint feature representation. The next dense layer followed by a dropout layer processes the concatenated features and prevents overfitting. A final dense layer with a single neuron and linear activation produces a scalar score representing the discriminator's assessment of how close the input channel sample is to the true channel distribution.

TABLE II: Dataset Parameters

Parameter	Description	Value
N_T	Number of BS transmit antennas	2
J	Number of UEs	2
T	Simulation time steps	2000
f	Carrier frequency	300 GHz
B	System bandwidth	1 GHz
N_F	Receiver noise figure	5 dB
$\mathcal{A}_{\mathcal{G}_t}$	Transmitter antenna gain	40 dBi
$\mathcal{A}_{\mathcal{G}_r}$	Receiver antenna gain	40 dBi
d_1	BS-RIS distance	2 m
$d_{2,j}$	Initial RIS-UE distances	$\{8, 10\}$ m
Λ_m	Amplitude reflection coefficient	1
θ_m	Phase shift range	$[0, 2\pi]$
v	UE speed per time step	0.1 m/step
ϱ	Molecular absorption coefficient	3.18×10^{-4}
Ω	Expected value of the channel matrix	0.22

The discriminator is trained using the CWGAN-GP loss function [93]. The discriminator loss is formulated as

$$\mathcal{L}_D = \mathcal{L}_W + \lambda_{GP} \cdot \mathcal{L}_{GP}, \tag{11}$$

where the Wasserstein distance component and the gradient penalty term defined as

$$\mathcal{L}_W = \mathbb{E}_{r\sim p_Y}\big[D(r,\hat{g})\big] - \mathbb{E}_{(r,g)\sim p_{Y,G}}\big[D(r,g)\big], \tag{12}$$

$$\mathcal{L}_{GP} = \mathbb{E}_{\tilde{g}\sim p_{\tilde{g}}}\left[\left(\|\nabla_{\tilde{g}} D(r,\tilde{g})\|_2 - 1\right)^2\right], \tag{13}$$

where $\tilde{g} = \epsilon g + (1-\epsilon)\,\hat{g}$ represents interpolated samples between the real channels g and the generated channels, with $\epsilon \sim \mathcal{U}[0,1]$. The coefficient λ_{GP} controls the strength of the gradient penalty, enforcing the 1-Lipschitz constraint on the discriminator. By combining the Wasserstein adversarial loss from eq. (12), the gradient penalty defined in eq. (13), and the L_2 reconstruction loss given in eq. (10), the overall training objective of the proposed TRiNCE model is summarized in eq. (14), which jointly optimizes the generator and discriminator through adversarial and reconstruction-based learning.

$$\begin{aligned}\min_G \max_D \mathcal{L} = {} & \mathbb{E}_{(r,g)\sim p_{Y,G}}[D(r,g)] - \mathbb{E}_{r\sim p_Y}[D(r,\hat{g})] \\ & - \lambda_{GP}\,\mathbb{E}_{\tilde{g}\sim p_{\tilde{g}}}\left[\left(\|\nabla_{\tilde{g}} D(r,\tilde{g})\|_2 - 1\right)^2\right] \\ & + \xi_{L2}\,\|g-\hat{g}\|_2^2.\end{aligned} \tag{14}$$

B. Simulation Setup

The dataset from Algorithm 1 is generated based on system parameters used in [40], [45], [71]. The system parameters used for dataset generation are summarized in Table II. The generator has two GRU layers. The first GRU layer consists of 64 units, and the second layer consists of 32 units, allowing the network to extract temporal features and capture both long-term and short-term channel features. The dense layer consists of 64 units. After the 50$^{\text{th}}$ epoch, α is decayed by 20% at every 10 epochs, while the learning rates for both the generator and discriminator are decayed by 10% every 15 epochs to ensure stable convergence. The proposed model parameters are listed in Table III, and the model is tested using the parameters specified in Table IV.

TABLE III: Proposed Model Parameters

Parameter	Description	Value
τ	Input Sequence	10
n_g	Generator updates	1
n_d	Discriminator updates	2
ξ_{L_2}	Reconstruction loss weight	100
B	Batch Size	128
E_p	Total Epochs	100
η_G	Initial learning rate of generator	1×10^{-3}
η_D	Initial learning rate of discriminator	1×10^{-3}

TABLE IV: Test Parameter Configurations

Parameter	Description	Tested Values
N_T	Number of BS antennas	$\{1, 2, 3, 4, 5\}$
γ	Transmit SNR	10 : 5: 40 dB
(α, μ)	Fading parameters	$\{(2,1), (2,2), (2,3), (3,2), (4,3.5), (5,4), (6,5.5)\}$
M	Number of RIS elements	$\{20, 25, 30, 35, 40\}$
ω_1	Power allocation factor for User 1	$\{0.3, 0.35, 0.4, 0.45, 0.5\}$

C. Inference FLOPs and Computational Complexity

For UE-side deployment, only the generator is utilized for inference. The computational cost is dominated by the two cascaded GRU layers with 64 and 32 hidden units over $\tau = 10$ time steps, resulting in approximately 253×10^3 and 184×10^3 FLOPs, respectively. The intermediate dense layer contributes about 4×10^3 FLOPs, while the final output layer scales linearly with the number of output channels $F_o = 2N_T$, contributing $130F_o$ FLOPs. Overall, the total inference complexity is approximately $441 \times 10^3 + 130F_o$ FLOPs.

The time complexity per inference is expressed as $\mathcal{O}(\tau(2H_1 + H_1^2) + \tau(H_1H_2 + H_2^2) + H_2D + DF_o)$, where τ denotes timesteps, H_1 and H_2 represent the dimensions of the cascaded GRU hidden states, D is the dense layer size, and F_o indicates the number of output channels. In the proposed deployment configuration, where τ, H_1, H_2, and D are fixed, the computational complexity simplifies to $\mathcal{O}(F_o)$, scaling linearly only with the output dimension.

D. Data Preprocessing

The raw dataset generated using Algorithm 1 consists of complex-valued data representing the received signal $\mathbf{Y}$ and cascaded channel coefficients $\{\mathbf{G}_{t,j,n_t}\}_{n_t=1}^{N_T}\}_{j=1,t=1}^{J,E}$. Since DL models cannot directly process complex-valued signals, the real and imaginary parts of $\mathbf{Y}$, and $\mathbf{G}$ are separated. In implementation the real and imaginary parts are saved as two separate columns named real and imag in CSV files. Separate CSV files are created for each user and each transmitter antenna. These CSV files are loaded and converted into real valued tensors. The received signal at the UE serves as the conditions for the generator network, providing it with the necessary information for accurate channel estimation.

The dataset is partitioned into training, validation, and testing sets using a 70:15:15% ratio. Since the received signal

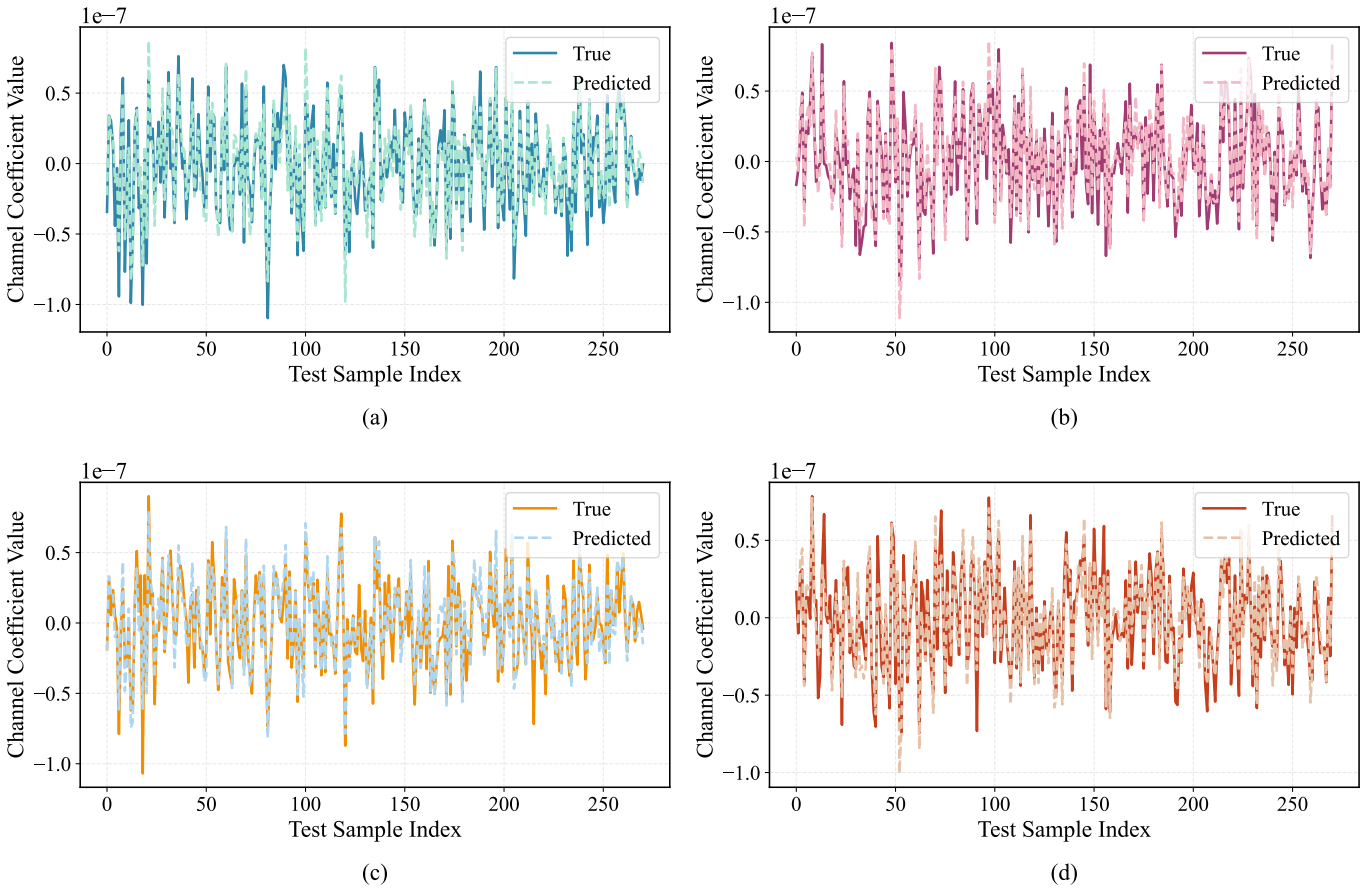


Fig. 5: Predicted and true channel coefficients on User 1 test data (a) G_0 real component, (b) G_0 imaginary component, (c) G_1 real component, and (d) G_1 imaginary component.

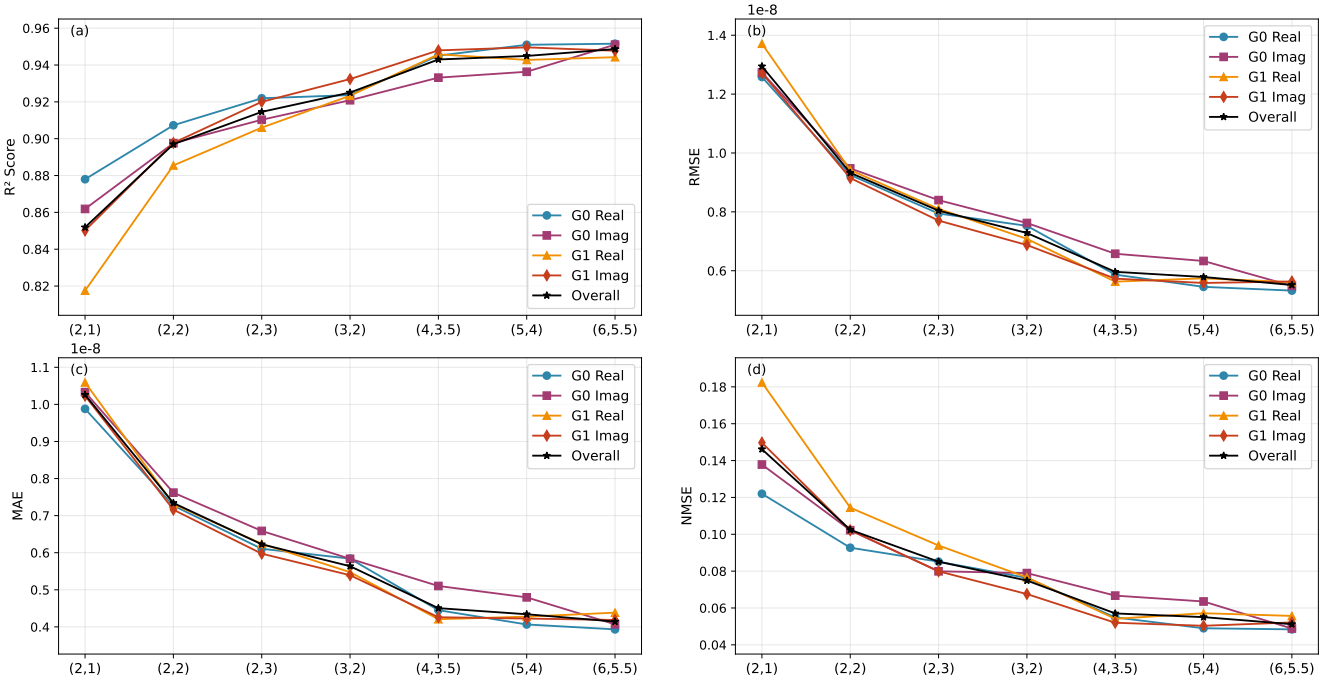


Fig. 6: Performance metrics for User 1 across different α–μ fading configurations, with power allocation coefficients of 0.3 for User 1 and 0.7 for User 2, and 20 RIS elements. (a) R^2 scores, (b) RMSE values, (c) MAE, and (d) NMSE.

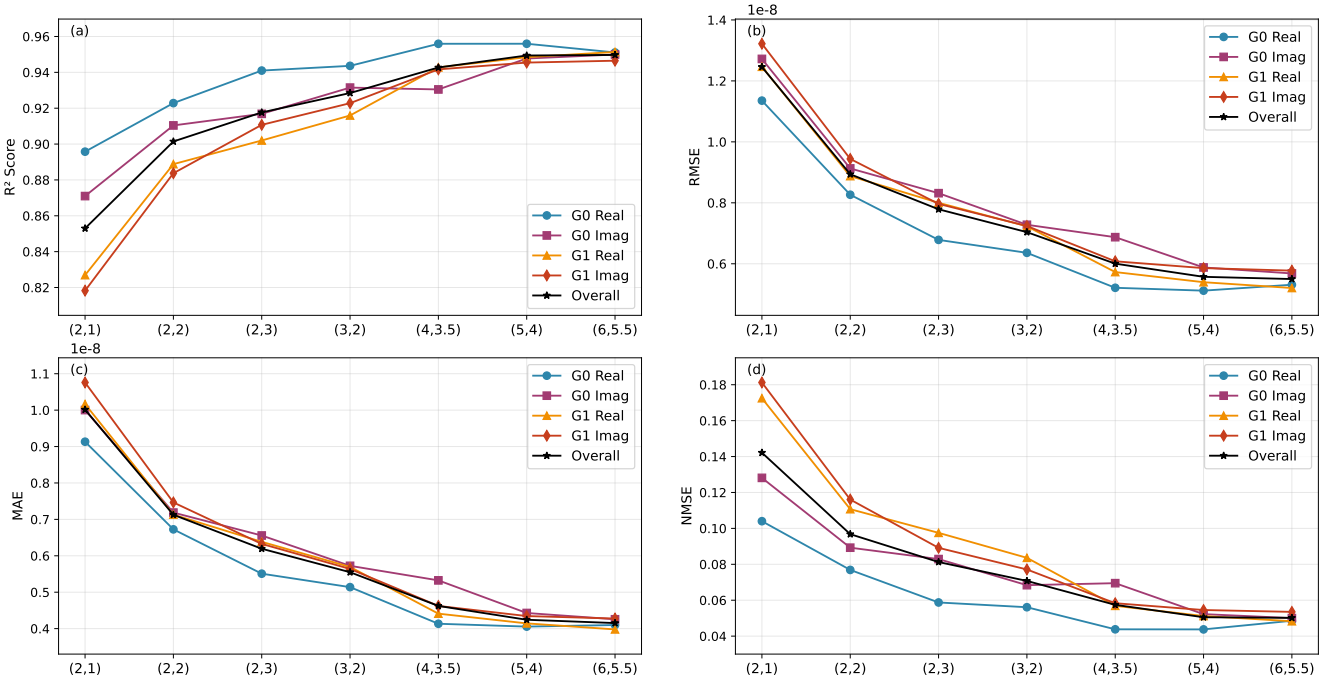


Fig. 7: Performance metrics for User 2 across different α–μ fading configurations, with power allocation coefficients of 0.3 for User 1 and 0.7 for User 2, and 20 RIS elements. (a) R^2 scores, (b) RMSE values, (c) MAE, and (d) NMSE.

and channel coefficients have different numerical ranges, standardization is applied to ensure that each feature has a zero mean and unit variance. The standardization parameters are computed only from the training set to prevent data leakage and maintain the integrity of the validation and test sets for unbiased performance evaluation.

A sliding window approach with window size τ is then applied to convert the time series into supervised learning samples, where each input sequence consists of τ consecutive observations of the received signal, and the corresponding cascaded channel coefficients serve as the target.

E. Evaluation Metrics

Once the generator is trained, the channel estimation performance of the TRiNCE framework is evaluated using root mean squared error (RMSE), mean absolute error (MAE), NMSE, and the coefficient of determination R^2. RMSE quantifies the average deviation of the estimated channel coefficients from the target channel coefficients, effectively penalizing larger errors. MAE provides a robust assessment of the average estimation accuracy. While RMSE and MAE provide valuable absolute error measurements, their scale-dependence makes it difficult to compare models across different datasets with varying scale. To address this, NMSE is a scale-independent metric that normalizes the estimation error power by the actual channel power, thereby indicating how well the model reconstructs the channel relative to the signal strength. Finally, the R^2 score evaluates how well the generated channel statistics match the variance of the actual wireless channel environment.

In terms of performance trends, lower values for the error metrics (RMSE, MAE, and NMSE) indicate higher estimation accuracy, whereas a higher R^2 score , closer to 1.0, indicates a better fit to the actual channel statistics. These metrics are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(g_i - \hat{g}_i)^2}, \tag{15}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|g_i - \hat{g}_i|, \tag{16}$$

$$\text{NMSE} = \frac{\sum_{i=1}^{N}|g_i - \hat{g}_i|^2}{\sum_{i=1}^{N}|g_i|^2}, \tag{17}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}|g_i - \hat{g}_i|^2}{\sum_{i=1}^{N}|g_i - \bar{g}|^2}, \tag{18}$$

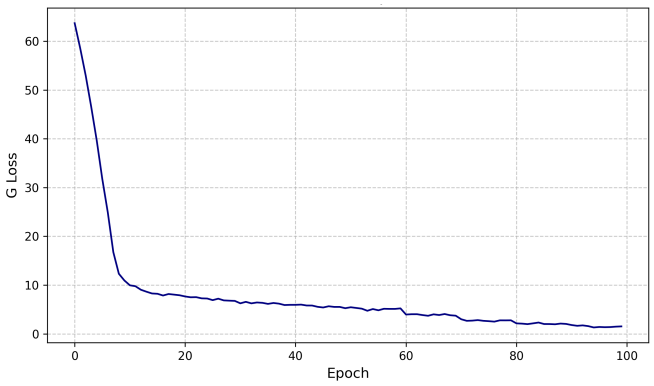
where N denotes the total number of test channel samples, g_i represents the actual target channel coefficient, $\hat{g}_i$ denotes the estimated channel coefficient, and $\bar{g} = \frac{1}{N}\sum_{i=1}^{N} g_i$ is the mean of the actual channel response.

V. Results and Discussion

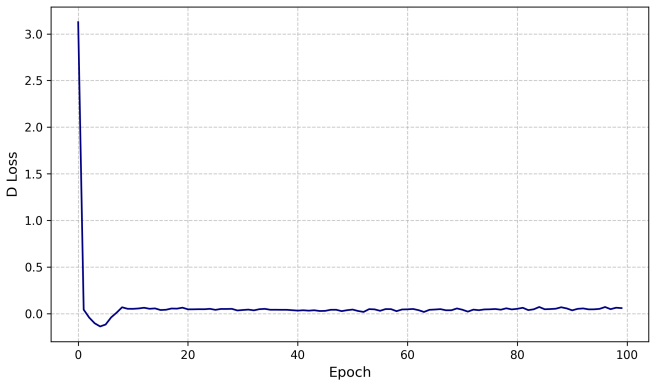
A. Model Performance Under a Specific RIS-NOMA Scenario

The performance of the TRiNCE model is evaluated using a testing configuration of α and μ values of $(2, 1)$, corresponding to the Rayleigh fading scenario, 20 RIS elements, a power allocation factor of 0.3, and two base station antennas. In this setup, G_0 represents the cascaded channel between the UE and BS antenna 0, and G_1 represents the cascaded channel between the UE and BS antenna 1. Each channel component is complex-valued, consisting of real and imaginary parts.

Fig. 8 illustrates the training loss history for both the generator and the discriminator over 100 epochs. As observed in Fig. 8a, the generator loss decreases rapidly during the initial epochs and then stabilizes, indicating that the model successfully learns the underlying channel distribution. Simultaneously, the discriminator loss as shown in Fig. 8b converges to a stable equilibrium near zero, demonstrating that the adversarial training process has reached a stable state without mode collapse or divergence.



(a) Generator Loss



(b) Discriminator Loss

Fig. 8: Training convergence behavior of the TRiNCE model showing (a) Generator loss and (b) Discriminator loss over epochs.

Fig. 5 shows the model's prediction performance on unseen test data. The time-series plots show that the predicted values closely track the true channel coefficients, with minimal deviation throughout the test sequence. The model achieves high prediction accuracy across all channel components. For User 1, the R^2 were 0.8719 and 0.8624 for the G_0 real and imaginary components, respectively. For G_1, the scores were 0.8325 (real) and 0.8466 (imaginary), resulting in an overall accuracy of 0.8533. Similarly, for User 2, the model attains R^2 scores of 0.8958 for G_0 real, 0.8710 for G_0 imaginary, 0.8269 for G_1 real, and 0.8182 for G_1 imaginary components, with an overall accuracy of 0.8530. A complete set of performance metrics for both users, including RMSE, MAE, and NMSE for all components, is summarized in Table V. The consistently high performance across both users, with overall accuracies exceeding 85%, demonstrates the model's effectiveness in multi-user RIS-NOMA channel prediction scenarios.

TABLE V: Performance Metrics Comparison for User 1 and User 2

Component	R^2	RMSE	MAE	NMSE
		User 1		
G_0 real	0.8780	1.26×10^{-8}	9.88×10^{-9}	0.1220
G_0 imag	0.8619	1.27×10^{-8}	1.03×10^{-8}	0.1378
G_1 real	0.8175	1.37×10^{-8}	1.06×10^{-8}	0.1824
G_1 imag	0.8502	1.27×10^{-8}	1.02×10^{-8}	0.1497
Overall	0.8519	1.29×10^{-8}	1.03×10^{-8}	0.1462
		User 2		
G_0 real	0.8958	1.14×10^{-8}	9.13×10^{-9}	0.1040
G_0 imag	0.8710	1.27×10^{-8}	1.00×10^{-8}	0.1281
G_1 real	0.8269	1.25×10^{-8}	1.02×10^{-8}	0.1725
G_1 imag	0.8182	1.32×10^{-8}	1.08×10^{-8}	0.1812
Overall	0.8530	1.25×10^{-8}	1.00×10^{-8}	0.1422

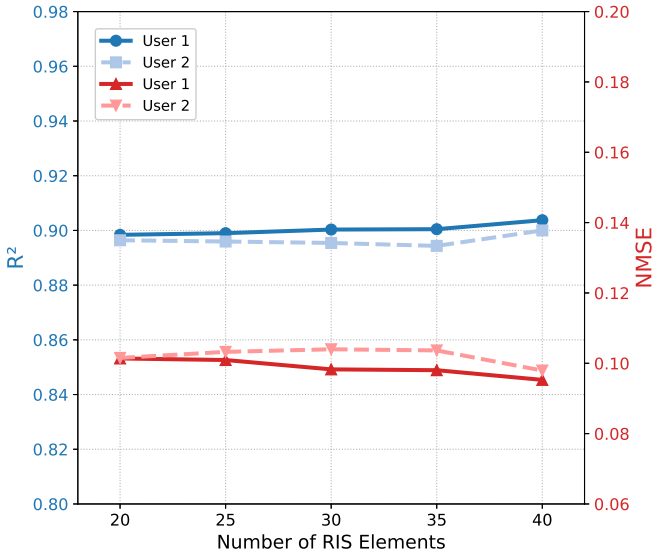


Fig. 9: Overall performance of the proposed channel estimation model under varying numbers of RIS elements.

B. Model Performance Under Varying Channel Conditions

Fig. 6 and Fig. 7 present the model's performance for User 1 and User 2 under different α–μ fading configurations. These configurations range from α and μ values of $(2, 1)$, which is for Rayleigh fading, to α and μ values of $(6, 5.5)$, which denotes milder fading conditions. As the severity of fading decreases, the model's prediction accuracy improves significantly. For User 1 this is reflected by overall R^2 scores increasing from 0.852 under Rayleigh fading to 0.949 under milder fading conditions. Error metrics also show a significant reduction of overall NMSE values reducing from 0.146 to 0.051. This trend highlights that under milder fading conditions, the channel becomes more predictable, enabling the model to learn the underlying nonlinear mapping between the received signal and the channel response more effectively. Furthermore, both real and imaginary components show similar performance behavior, confirming that the model effectively

TABLE VI: Performance and Complexity Comparison of Different Models

Model	RMSE	R^2	MAE	NMSE	Parameters
CNN	1.15×10^{-8}	0.8433	9.14×10^{-9}	0.1555	223,812
CNN-LSTM	1.03×10^{-8}	0.8756	8.12×10^{-9}	0.1243	650,372
GRU	1.03×10^{-8}	0.8757	8.26×10^{-9}	0.1238	64,452
LSTM	1.02×10^{-8}	0.8772	8.15×10^{-9}	0.1227	77,444
Proposed Model	$\mathbf{9.16 \times 10^{-9}}$	**0.9006**	$\mathbf{7.24 \times 10^{-9}}$	**0.0988**	**53,893**

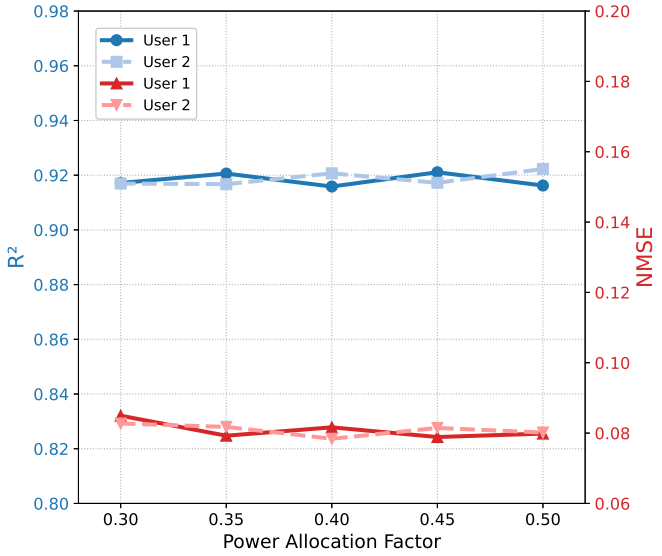


Fig. 10: Overall performance of the proposed channel estimation model under varying power allocation factor w_1.

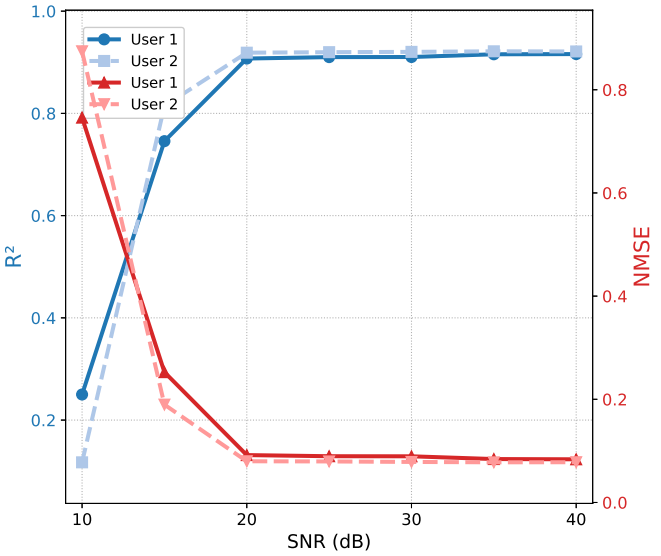


Fig. 12: Overall performance of the proposed channel estimation model under Different SNR conditions.

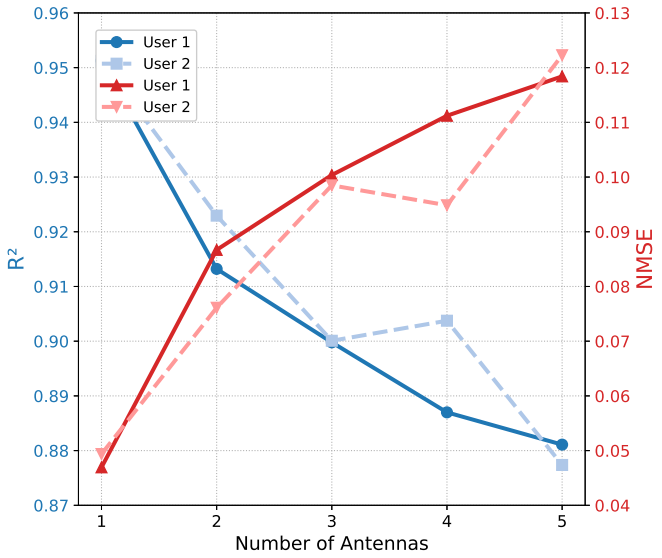


Fig. 11: Overall performance of the proposed channel estimation model for different numbers of BS antennas.

handles the complex-valued nature of the signal. Similar results are observed for User 2, where overall R^2 score improves from 0.853 to 0.950 and overall NMSE decreases from 0.142 to 0.050, showing that the model is robust, stable, and has good generalization capability across different users and fading environments. Overall, these results show that the model is effective in accurately estimating channel characteristics under a wide range of channel conditions.

C. Model Performance Under Varying Number of RIS Elements

The performance of the proposed channel estimation model is evaluated for different numbers of RIS elements with α and μ values of $(2, 2)$ and $\omega_1 = 0.3$. For User 1, the R^2 scores of the real and imaginary parts of the cascaded channels (G_0 and G_1) remain consistently high across all RIS sizes, with overall R^2 values ranging from 0.898 to 0.904. Similarly, the NMSE values are low and stable, ranging from 0.095 to 0.111, indicating accurate channel estimation. For User 2, the R^2 scores and NMSE values show similar trend of staying low and stable for different number of RIS elements.
Overall, the variation across all RIS settings is minimal, with a maximum change of only 1.1% in R^2, indicating that the model remains robust and stable across all RIS configurations for both users. This minimal variation demonstrates the proposed model's scalability. Despite the increased dimensionality and complexity of the channel associated with larger RIS surfaces, the proposed framework maintains reliable estimation performance. These results are shown in Fig. 9.

D. Model Performance Under Varying Power Allocation Factor

In Fig. 10, the overall performance of the proposed model can be observed for different power allocation factors w_1 ranging from 0.3 to 0.5 with α and μ values of $(2, 3)$, 20 RIS elements, and two BS antennas. The overall R^2 values for the real and imaginary components of the G_0 and G_1 channels consistently exceed 0.9 for both users, achieving the highest value of 0.922 for User 2 at $w_1 = 0.5$.

For User 1, the real and imaginary components of G_0 have high R^2 values, ranging from 0.9145 to 0.9240 for the real component and 0.9159 to 0.9257 for the imaginary component. The minimum variation suggests that the proposed model maintains stable performance for different NOMA power allocation factors. Similar stability can be observed for channel G_1, both in the real and imaginary parts, with values ranging from 0.907 to 0.925. A similar trend is observed for User 2. The R^2 values for the real and imaginary components of G_0 vary between 0.9362 and 0.9418 and 0.9096 and 0.9249, respectively. The G_1 components have slightly lower R^2 values, but they are still consistently high. The corresponding NMSE values show a declining trend, in agreement with the higher R^2 values. For User 1, the NMSE values for G_0 range from 0.077 to 0.097, while for G_1, they vary between 0.075 and 0.087. The NMSE values are slightly lower for User 2, ranging between 0.058 and 0.095 across all components.

Overall, this stability of R^2 and NMSE values across varying power allocation factors is significant for NOMA implementations, where power allocation factors are dynamically adjusted to meet user fairness conditions. The results indicate that the proposed model is resilient to these dynamic adjustments.

E. Model Performance with Different Numbers of BS Antennas

The performance of the proposed channel estimation model is evaluated for different numbers of transmit antennas (N_t) at the base station. Fig. 11 shows the performance averaged across the real and imaginary channel coefficients of all transmit antennas, N_t. For User 1, the R^2 value decreases slightly from 0.9513 for $N_t = 1$ to 0.8811 for $N_t = 5$, and the NMSE increases from 0.0469 to 0.1184, showing some reduction in estimation accuracy with more antennas. User 2 shows a similar trend, with R^2 decreasing from 0.9512 to 0.8774 and NMSE increasing from 0.0494 to 0.1223. These results demonstrate the scalability of the proposed model, which maintains robust performance even as the number of base station antennas increases, with a performance degradation of approximately 7.38% for User 1 and 7.76% for User 2. This degradation is because of the increased dimensionality of the output, as the model must simultaneously predict $N_t \times 2$ values, corresponding to the real and imaginary components for each channel path, increasing the complexity of the regression task.

F. Model Performance under Different SNR Conditions

The TRiNCE performance under varying SNR conditions is shown in Fig. 12. At low SNRs, the channel estimation is dominated by noise. At 15 dB SNR, the model already demonstrates a strong capability to distinguish signal features from the noise floor. As SNR increases to 20 dB, the estimation accuracy improves, with User 2 achieving an R^2 of 0.918 and User 1 achieving 0.907. The NMSE correspondingly drops below 0.1 for both users. Beyond 20dB, a performance plateau is reached where increasing the SNR from 20 dB to 40 dB yields negligible improvement and the NMSE stabilizes around 0.08 and the R^2 remains constant at approximately 0.92. This shows that the proposed framework is highly robust, achieving near optimal performance even at moderate SNR levels typical of practical deployment.

G. Comparison of the Proposed Model With Other DL Models

The performance of the proposed model was compared with other DL architectures widely used in literature for time series prediction tasks. These models include 1D-CNN, LSTM, GRU, and a 1D-CNN-LSTM hybrid model. The 1D-CNN captures local temporal features, recurrent models such as LSTM and GRU are effective in modeling long-term dependencies, and the hybrid 1D-CNN-LSTM leverages the strengths of both convolutional and recurrent structures, making it a strong benchmark model. The Adam optimizer with a learning rate of 10^{-3} and mean squared error (MSE) as the loss function was used in the training of each model. And the batch size is set to 128. The baseline model architectures are discussed as follows.

1) CNN Model: The CNN model uses three sequential 1D convolutional layers with 64, 128, and 256 filters, each having a kernel size of 3 and ReLU activation. This is followed by batch normalization to stabilize the learning process. A global average pooling layer is then applied to average the temporal features across time. Finally, two dense layers with 256 and 128 units and ReLU activations, along with dropout rates of 0.3 and 0.2, are used before the final dense layer with F_o units.

2) LSTM Model: The LSTM model consists of four LSTM layers, with 64, 64, 32 and 32 units, respectively. The first three layers return sequences. The LSTM outputs are given to two dense layers with 64 units. A dropout layer with a rate of 0.15 is applied after each dense layer. The dense layers use ReLU activation functions and linear output layer is of F_o units.

3) GRU Model: The GRU model has the same structure as the LSTM model with GRU layers instead.

4) CNN–LSTM Model: The CNN–LSTM model has convolutional and recurrent layers. The CNN block has two Conv1D layers with 64 and 128 filters, kernel size 3, ReLU activation, and batch normalization. The extracted features are then passed to two LSTM layers with 128 and 256 units for temporal processing. The dense block comprises two fully connected layers with 256 and 128 neurons, dropout rates of 0.3 and 0.2, and a final linear output layer with F_O units.

As shown in Table VI, the proposed model achieves the lowest RMSE, MAE, and NMSE values while achieving the highest R^2 among all compared architectures. Moreover, the proposed model uses significantly fewer trainable parameters compared to CNN, LSTM, GRU, and CNN–LSTM models, which have hundreds of thousands of parameters. These results

indicate that the proposed model not only provides more accurate and reliable channel estimation but also offers a more computationally efficient solution.

VI. Conclusion and Future Work

This paper has addressed the challenge of channel estimation by proposing TRiNCE, a DL framework, utilizing a GRU-based cWGAN-GP. The GRU-based architecture effectively captures the sequential and temporal patterns present in the received signal sequence. Furthermore, the cWGAN-GP architecture learns and produces channel coefficients that closely follow the statistical distribution of the actual physical channel. We have simulated a downlink MISO system and comprehensively evaluated the performance of TRiNCE models against several key system parameters, including varying channel severity by adjusting the $\alpha - \mu$ values, increasing the number of RIS elements, power allocation factors, and the number of antennas. The results confirmed that as the fading severity decreases, the channel estimation performance improves. The model maintained high performance with minimal deviation as the number of RIS elements and power allocation factors increased. While the model's performance slightly degraded with the increase in the number of BS antennas, the proposed model's estimation accuracy remained high across MISO settings. These results demonstrate the stability and robustness of the TRiNCE model in different RIS-NOMA configurations under varying channel conditions in MISO scenarios. In comparison to the baseline DL models, including 1D-CNN, LSTM, GRU, and 1D-CNN-LSTM hybrid networks, TRiNCE demonstrated superior accuracy and robustness with fewer trainable parameters, establishing it as a computationally efficient and accurate channel estimation model. To facilitate real-world implementation, the proposed architecture decouples the computationally intensive training from the online inference phase. Although the training of the GAN architecture is computationally intensive, this complexity is strictly confined to the offline phase. The online system utilizes only the lightweight generator, ensuring the low latency required for THz communications. Moreover, this lightweight design facilitates adaptability in dynamic environments. Having been validated across generalized $\alpha-\mu$ fading conditions, the proposed model is proven to handle a diverse range of channel statistics ensuring reliable operation across varying deployment scenarios.

In future work, we can analyze and adapt the TRiNCE framework to perform robust channel estimation in the presence of realistic hardware impairments, such as non-linear components and THz-specific phase noise. Additionally, while this study assumes perfect CSI at the BS to establish a performance baseline and employs SVD-based beamforming under this assumption, a closed-loop scenario where the TRiNCE-estimated channel is utilized by the BS for optimized BS beamforming and power allocation can be considered. Finally, we can explore the feasibility of a QML, e.g., a quantum gated recurrent unit (Q-GRU), to implement the TRiNCE model. This direction could potentially offer exponential speedup in channel estimation complexity and significantly improve resource efficiency, offering a solution for future quantum-enabled wireless networks.

Acknowledgments

This work was supported in part by the Canada Excellence Research Chair (CERC) Program CERC-2022-00109 and in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant Program RGPIN-2025-04941. The work of B. Canberk is partially supported by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) 1515 Frontier R&D Laboratories Support Program for Türk Telekom 6G R&D Lab under project number 5249902.

References

[1] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[2] X. Li, X. Gao, L. Yang, H. Liu, J. Wang, and K. M. Rabie, "Performance analysis of STAR-RIS-CR-NOMA-based consumer IoT networks for resilient industry 5.0," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 1380–1391, Sep. 2024.

[3] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: Vision and potential techniques," *IEEE Netw.*, vol. 33, no. 4, pp. 70–75, Jul. 2019.

[4] A. Masaracchia, D. van Huynh, T. Q. Duong, O. A. Dobre, A. Nallanathan, and B. Canberk, "The role of digital twin in 6G-based URLLCs: Current contributions, research challenges, and next directions," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 1202–1215, Feb. 2025.

[5] D. Sarkarand, S. S. Yadav, V. Pal, K. Singh, K. Dev, and H. Shin, "A multi-metric performance analysis of STAR-RIS aided MIMO-NOMA framework for ciot-enabled 6G wireless networks," *IEEE Trans. Consum. Electron.*, pp. 1–1, Dec. 2025.

[6] T. Bui, A. Masaracchia, V. Sharma, O. Dobre, and T. Duong, "Impact of 6G space-air-ground integrated networks on hard-to-reach areas: Tourism, agriculture, education, and indigenous communities," *EAI Endorsed Tour Tech Intel*, vol. 1, no. 1, Sep. 2024.

[7] M. M. Kiasaraei, K. Nikitopoulos, and R. Tafazolli, "Toward ultra-power-efficient, tbps wireless systems via analogue processing: Existing approaches, challenges and way forward," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 2, pp. 747–780, Second Quarter 2024.

[8] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 ghz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, Jun. 2019.

[9] A. Shafie, N. Yang, C. Han, J. M. Jornet, M. Juntti, and T. Kürner, "Terahertz communications for 6G and beyond wireless networks: Challenges, key advancements, and opportunities," *IEEE Netw.*, vol. 37, no. 3, pp. 162–169, May/Jun. 2023.

[10] W. K. Alsaedi, H. Ahmadi, Z. Khan, and D. Grace, "Spectrum options and allocations for 6G: A regulatory and standardization review," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1787–1812, Aug. 2023.

[11] C. Huang, Z. Yang, G. C. Alexandropoulos, K. Xiong, L. Wei, C. Yuen, Z. Zhang, and M. Debbah, "Multi-hop RIS-empowered terahertz communications: A DRL-based hybrid beamforming design," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1663–1677, Jun. 2021.

[12] N. U. Saqib, M. S. Haroon, H. Y. Lee, K. Park, H.-G. Song, and S.-W. Jeon, "THz communications: A key enabler for future cellular networks," *IEEE Access*, vol. 11, pp. 117 474–117 493, Jan. 2023.

[13] D. Serghiou, M. Khalily, T. W. C. Brown, and R. Tafazolli, "Terahertz channel propagation phenomena, measurement techniques and modeling for 6G wireless communication applications: A survey, open challenges and future research directions," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 1957–1996, Fourth Quarter 2022.

[14] C. Han, Y. Wang, Y. Li, Y. Chen, N. A. Abbasi, T. Kürner, and A. F. Molisch, "Terahertz wireless channels: A holistic survey on measurement, modeling, and analysis," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1670–1707, Third Quarter 2022.

[15] S. Li, Y. Wu, Y. Zhang, S. Duan, J. Xu, and C. Li, "Privacy transmission via joint active and passive beamforming optimization for RIS-aided NOMA-IoMT networks," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 2290–2302, Jan. 2024.

[16] Z. Ali, M. Asif, W. U. Khan, A. Elfikky, A. Ihsan, M. Ahmed, A. Ranjha, and G. Srivastava, "Hybrid optimization for NOMA-based transmissive-RIS mounted UAV networks," *IEEE Trans. Consum. Electron.*, vol. 71, no. 2, pp. 3740–3752, Jan. 2025.

[17] D.-B. Ha, V.-T. Truong, and Y. Lee, "Intelligent reflecting surface assisted RF energy harvesting mobile edge computing NOMA networks: Performance analysis and optimization." *EAI Endorsed Trans. Ind. Networks Intell. Syst.*, vol. 9, no. 32, p. e5, Aug. 2022.

[18] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[19] P. Q. Truong, T. Do-Duy, V.-C. Phan, and A. Masaracchia, "Jointly power allocation and phase shift optimization for RIS empowered downlink cellular networks," *EAI Endorsed Trans Ind Net Intel Syst*, vol. 10, no. 4, p. e4, Dec. 2023.

[20] E. Basar, G. C. Alexandropoulos, Y. Liu, Q. Wu, S. Jin, C. Yuen, O. A. Dobre, and R. Schober, "Reconfigurable intelligent surfaces for 6G: Emerging hardware architectures, applications, and open challenges," *IEEE Veh. Technol. Mag.*, vol. 19, no. 3, pp. 27–47, Sep. 2024.

[21] F. Verde, D. Darsena, and V. Galdi, "Rapidly time-varying reconfigurable intelligent surfaces for downlink multiuser transmissions," *IEEE Trans. Commun.*, vol. 72, no. 6, pp. 3227–3243, Jun. 2024.

[22] M. Ahmed, A. Wahid, W. U. Khan, F. Khan, A. Ihsan, Z. Ali, K. M. Rabie, T. Shongwe, and Z. Han, "A survey on RIS advances in terahertz communications: Emerging paradigms and research frontiers," *IEEE Access*, vol. 12, pp. 173 867–173 901, Oct. 2024.

[23] T. Do-Duy, A. Masaracchia, B. Canberk, L. D. Nguyen, and T. Q. Duong, "Throughput maximization in RIS-assisted NOMA-THz communication network," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 5706–5717, Sep. 2024.

[24] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, M. Di Renzo, and N. Al-Dhahir, "Reconfigurable intelligent surfaces: Principles and opportunities," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1546–1577, May 2021.

[25] M. Deng, M. Ahmed, A. Wahid, A. A. Soofi, W. U. Khan, F. Xu, M. Asif, and Z. Han, "Reconfigurable intelligent surfaces enabled vehicular communications: A comprehensive survey of recent advances and future challenges," *IEEE Trans. Intell. Veh.*, pp. 1–28, Aug. 2024.

[26] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[27] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7244–7257, Nov. 2016.

[28] H. Q. Tran and M. D. Ong, "Performance analysis of NOMA over n-m fading channels with imperfect SIC," *EAI Endorsed Trans. Ind. Netw. Intell. Syst.*, vol. 10, no. 1, p. e5, May 2023.

[29] V. R. J. Velez, J. P. C. B. B. Pavia, N. M. B. Souto, P. J. A. Sebastião, and A. M. C. Correia, "Performance assessment of a RIS-empowered post-5G/6G network operating at the mmWave/THz bands," *IEEE Access*, vol. 11, pp. 49 625–49 638, May 2023.

[30] M. H. Kumar, S. Sharma, K. Deka, and M. K. Sharma, "RIS-assisted user pairing NOMA system for THz communications," in *Proc. Nat. Conf. Commun. (NCC)*, Guwahati, India, Feb. 2023, pp. 1–6.

[31] M. H. Kumar, S. Sharma, Y. Yoganandam, K. Deka, and A. Kherani, "Reconfigurable intelligent surfaces assisted SM-cooperative NOMA for THz communications," in *Proc. Nat. Conf. Commun. (NCC)*, Chennai, India, Feb. 2024, pp. 1–6.

[32] W. Luo, W. Liang, S. Wen, D. Wang, and L. Li, "Power allocation and passive beamforming design for RIS aided NOMA in THz communications," in *Proc. 10th Int. Conf. Inf. Syst. Comput. Technol. (ISCTech)*, Guilin, China, Dec. 2022, pp. 7–13.

[33] X. Xu, Q. Chen, X. Mu, Y. Liu, and H. Jiang, "Graph-embedded multi-agent learning for smart reconfigurable THz MIMO-NOMA networks," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 259–275, Jan. 2022.

[34] Y. Liu, X. Mu, X. Liu, M. Di Renzo, Z. Ding, and R. Schober, "Reconfigurable intelligent surface-aided multi-user networks: Interplay between NOMA and RIS," *IEEE Wireless Commun. Mag.*, vol. 29, no. 2, pp. 169–176, Apr. 2022.

[35] F. Fang, B. Wu, S. Fu, Z. Ding, and X. Wang, "Energy-efficient design of STAR-RIS aided MIMO-NOMA networks," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 498–511, Jan. 2023.

[36] J. Kokkoniemi, J. Lehtomäki, and M. Juntti, "Simple molecular absorption loss model for 200–450 gigahertz frequency band," in *Proc. 2019 Eur. Conf. Net. Commun. (EuCNC).*, Valencia, Spain, Jun. 2019, pp. 219–223.

[37] E. N. Papasotiriou, A.-A. A. Boulogeorgos, M. Francis De Guzman, K. Haneda, and A. Alexiou, "A new look to THz wireless links: Fading modeling and capacity assessment," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, Helsinki, Finland, Sep. 2021, pp. 1–5.

[38] M. D. Yacoub, "The α-μ distribution: A physical fading model for the stacy distribution," *IEEE Trans. Veh. Technol.*, vol. 56, no. 1, pp. 27–34, Jan. 2007.

[39] E. N. Papasotiriou, A.-A. A. Boulogeorgos, K. Haneda, M. F. de Guzman, and A. Alexiou, "An experimentally validated fading model for THz wireless systems," *Scientific Reports*, vol. 11, no. 1, p. 18717, Sep. 2021.

[40] V. K. Chapala and S. M. Zafaruddin, "Exact analysis of RIS-aided THz wireless systems over $\alpha-\mu$ fading with pointing errors," *IEEE Commun. Lett.*, vol. 25, no. 11, pp. 3508–3512, Nov. 2021.

[41] P. Bhardwaj and S. M. Zafaruddin, "Performance analysis of cooperative relaying for multi-antenna RF transmissions over THz wireless link," in *Proc. IEEE 95th Veh. Technol. Conf. (VTC Spring)*, Helsinki, Finland, Jun. 2022, pp. 1–5.

[42] S. Li and L. Yang, "Performance analysis of dual-hop THz transmission systems over $\alpha-\mu$ fading channels with pointing errors," *IEEE Internet of Things J.*, vol. 9, no. 14, pp. 11 772–11 783, Jul. 2022.

[43] A. Alqahtani, E. Alsusa, A. Al-Dweik, and M. Al-Jarrah, "Performance analysis for downlink NOMA over α-μ generalized fading channels," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6814–6825, Jul. 2021.

[44] Q. Li, M. El-Hajjar, Y. Sun, I. Hemadeh, A. Shojaeifard, Y. Liu, and L. Hanzo, "Achievable rate analysis of the STAR-RIS-aided NOMA uplink in the face of imperfect CSI and hardware impairments," *IEEE Trans. Commun.*, vol. 71, no. 10, pp. 6100–6114, Oct. 2023.

[45] C. Nguyen, T. M. Hoang, and A. A. Cheema, "Channel estimation using CNN-LSTM in RIS-NOMA assisted 6G network," *IEEE Trans. Mach. Learn. Commun. Netw.*, vol. 1, pp. 43–60, May 2023.

[46] N. Q. T. Thoong, A. A. Cheema, B. Canberk, D. T. Tran, O. A. Dobre, and T. Q. Duong, "Channel estimation for reconfigurable intelligent surface-aided 6G NOMA systems: A quantum machine learning approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 13, pp. 2197–2218, Sep. 2025.

[47] M. A. Jayasinghe, D. H. Gunasinghe, and G. A. A. Baduge, "Achievable rate analysis for RIS-aided massive MIMO NOMA systems with statistical composite channel knowledge," *IEEE Trans. Veh. Technol.*, vol. 73, no. 10, pp. 14 888–14 907, Oct. 2024.

[48] M. Chakraborty, E. Sharma, H. A. Suraweera, and H. Quoc Ngo, "Analysis and optimization of RIS-assisted cell-free massive MIMO NOMA systems," *IEEE Trans. Commun.*, vol. 73, no. 4, pp. 2631–2647, Apr. 2025.

[49] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4144–4157, Jun. 2021.

[50] S. Bhattacharya and A. K. Gupta, "Deep learning for THz channel estimation and beamforming prediction via sub-6GHz channel," in *Proc. IEEE Int. Conf. Signal Process. Commun. (SPCOM)*, Bangalore, India, Jul. 2022, pp. 1–5.

[51] P. Tarafder, I. Ahmed, D. B. Rawat, R. Annavajjala, and K. V. Mishra, "Deep learning model-based channel estimation for THz band massive MIMO with RF impairments," in *Proc. MILCOM - IEEE Mil. Commun. Conf. (MILCOM)*, Washington, DC, USA, Oct. 2024, pp. 282–287.

[52] C. Hongyun, P. Xue, and L. Baijiang, "Channel estimation for reconfigurable intelligent surface aided multiuser millimeter-wave/THz systems," *China Communications*, vol. 21, no. 3, pp. 91–103, Mar. 2024.

[53] J. Wu, S. Kim, and B. Shim, "Parametric sparse channel estimation for RIS-assisted terahertz systems," *IEEE Trans. Commun.*, vol. 71, no. 9, pp. 5503–5518, Sep. 2023.

[54] F. E. Asim, B. Sokal, A. L. F. de Almeida, B. Makki, and G. Fodor, "Structured channel estimation for RIS-assisted THz communications," *IEEE Trans. Veh. Technol.*, vol. 74, no. 3, pp. 5175–5180, Mar. 2025.

[55] Z. Li, Z. Chen, X. Ma, and W. Chen, "Channel estimation for intelligent reflecting surface enabled terahertz MIMO systems: A deep learning perspective," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC Workshop)*, Chongqing, China, Aug. 2020, pp. 75–79.

[56] Z. Wan, Z. Gao, F. Gao, M. D. Renzo, and M.-S. Alouini, "Terahertz massive MIMO with holographic reconfigurable intelligent surfaces," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4732–4750, Jul. 2021.

[57] J. Wang, C.-X. Wang, J. Huang, and Y. Chen, "6g THz propagation channel characteristics and modeling: Recent developments and future challenges," *IEEE Commun. Mag.*, vol. 62, no. 2, pp. 56–62, Feb. 2024.

[58] J. Kokkoniemi, J. Lehtomäki, and M. Juntti, "A discussion on molecular absorption noise in the terahertz band," *Nano communication networks*, vol. 8, pp. 35–45, Jun. 2016.

[59] J. M. Jornet and I. F. Akyildiz, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3211–3221, Oct. 2011.

[60] N. Elburki, S. Ben Amor, and S. Affes, "Evaluation of path-loss models for THz propagation in indoor environments," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, London, ON, Canada, Aug. 2020, pp. 1–5.

[61] J. Kokkoniemi, J. Lehtomäki, and M. Juntti, "Simplified molecular absorption loss model for 275–400 gigahertz frequency band," in *Proc. 12th Eur. Conf. Antennas Propag. (EuCAP)*, London, UK,, Apr. 2018, pp. 1–5.

[62] H. Du, J. Zhang, K. Guan, D. Niyato, H. Jiao, Z. Wang, and T. Kürner, "Performance and optimization of reconfigurable intelligent surface aided THz communications," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3575–3593, May 2022.

[63] M. Erdem, A. Saleem, O. Gurbuz, C. Ecemis, and A. Saeed, "A simple analytical model for THz band path loss," *IEEE Commun. Lett.*, vol. 27, no. 3, pp. 996–1000, Mar. 2023.

[64] K. Guan, H. Yi, D. He, B. Ai, and Z. Zhong, "Towards 6G: Paradigm of realistic terahertz channel modeling," *China Communications*, vol. 18, no. 5, pp. 1–18, May 2021.

[65] K. Zhang, H. Wang, C. Zhang, X. Yu, and Y. Dong, "Three-dimensional geometry-based stochastic model for sub-terahertz air-to-air UAV-MIMO channels," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Dalian, China, Aug. 2023, pp. 1–6.

[66] Y. Yuan, R. He, B. Ai, Z. Ma, Y. Miao, Y. Niu, J. Zhang, R. Chen, and Z. Zhong, "A 3D geometry-based THz channel model for 6G ultra massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 2251–2266, Mar. 2022.

[67] A. R. Ekti, A. Boyaci, A. Alparslan, Ünal, S. Yarkan, A. Görçin, H. Arslan, and M. Uysal, "Statistical modeling of propagation channels for terahertz band," in *Proc. IEEE Conf. Stand. Commun. Netw. (CSCN)*, Helsinki, Finland, Sep. 2017, pp. 275–280.

[68] J. M. Romero-Jerez, F. J. Lopez-Martinez, J. F. Paris, and A. J. Goldsmith, "The fluctuating two-ray fading model: Statistical characterization and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4420–4432, Jul. 2017.

[69] H. Lin, Y. Li, G. Chen, Z. Liu, and Y. Huang, "Performance analysis for mmwave cell-free access network based on terahertz backhaul," *IEEE Commun. Lett.*, vol. 29, no. 6, pp. 1181–1185, Jun. 2025.

[70] M. López-Benítez and M. Alhulayil, "On the computational efficiency of fading models: The fluctuating two-ray model for mmWave bands," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 8175–8189, Sep. 2025.

[71] N. P. Le and M.-S. Alouini, "Performance analysis of RIS-aided THz wireless systems over $\alpha - \mu$ fading: An approximate closed-form approach," *IEEE Internet of Things J.*, vol. 11, no. 1, pp. 1328–1343, Jan. 2024.

[72] ——, "Performance analysis of STAR-RIS-assisted NOMA wireless systems with realistic indoor outdoor THz channel models," *IEEE Trans. Veh. Technol.*, pp. 1–16, Aug. 2025.

[73] V. Kondepogu and B. Bhattacharyya, "Heuristic optimization-assisted dilated convolution neural network with gated recurrent unit for channel estimation in NOMA-OFDM system," *IEEE Access*, vol. 12, pp. 184 456–184 476, Oct. 2024.

[74] S. Khaled, D. Abdelhameed, M. Hassan Essai Ali, H. Ghazi Zaini, S. Elnazer, and S. Heshmat, "NOMA channel state estimation: Deep learning approaches," *IEEE Access*, vol. 13, pp. 159 774–159 795, Sep. 2025.

[75] K. Dovelos, M. Matthaiou, H. Q. Ngo, and B. Bellalta, "Channel estimation and hybrid combining for wideband terahertz massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1604–1620, Jun. 2021.

[76] Z. Lin, L. Wang, J. Ding, B. Tan, and S. Jin, "Channel power gain estimation for terahertz vehicle-to-infrastructure networks," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 155–159, Jan. 2023.

[77] Y. Li and A. S. Madhukumar, "Hybrid near- and far-field thz ummimo channel estimation: A sparsifying matrix learning-aided bayesian approach," *IEEE Trans. Wireless Commun.*, vol. 24, no. 3, pp. 1881–1897, Mar. 2025.

[78] Y. Dai, W. Zhao, Y. You, Y. Huang, and C. Zhang, "Local subdivisional channel estimation in RIS-assisted THz MIMO-OFDM systems," in *Proc. 5th Inf. Commun. Technol. Conf. (ICTC).*, Nanjing,China, May 2024, pp. 234–238.

[79] L. Chen, Z. Tan, and J. Cao, "Deep learning-based terahertz channel estimation," in *Proc. Cross Strait Radio Sci. Wireless Technol. Conf. (CSRSWTC)*, Haidian, China, Dec. 2022, pp. 1–4.

[80] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[81] Q. Spencer, C. Peel, A. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 60–67, Oct. 2004.

[82] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[83] B. Clerckx, Y. Mao, R. Schober, E. A. Jorswieck, D. J. Love, J. Yuan, L. Hanzo, G. Y. Li, E. G. Larsson, and G. Caire, "Is NOMA efficient in multi-antenna networks? a critical look at next generation multiple access techniques," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1310–1343, Jun. 2021.

[84] T. Hou, Y. Liu, Z. Song, X. Sun, and Y. Chen, "MIMO-NOMA networks relying on reconfigurable intelligent surface: A signal cancellation-based design," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 6932–6944, Nov. 2020.

[85] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Adv. Neural Inf. Process*, vol. 27, Jun. 2014.

[86] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, Nov. 2014.

[87] F. Han, Y. Liang, Q. Ling, and H. Han, "An improved conditional wasserstein gan with gradient penalty for gene expression profiling data augmentation based on data segmentation and depth feature constraint," *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, vol. 22, no. 4, pp. 1401–1414, Jul.-Aug. 2025.

[88] M. Zheng, T. Li, R. Zhu, Y. Tang, M. Tang, L. Lin, and Z. Ma, "Conditional wasserstein generative adversarial network-gradient penalty-based approach to alleviating imbalanced data classification," *Information Sciences*, vol. 512, pp. 1009–1023, Feb. 2020.

[89] K. Cho, B. van Merrienboer, Çaglar Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, Doha, Qatar, Oct. 2014, pp. 1–15.

[90] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, Dec. 2014.

[91] P. Ramachandran, B. Zoph, and Q. V. Le, "Swish: a self-gated activation function," *arXiv: Neural and Evolutionary Computing*, Oct. 2017.

[92] Y. Li, N. Xiao, and W. Ouyang, "Improved generative adversarial networks with reconstruction loss," *Neurocomputing*, vol. 323, pp. 363–372, Jan. 2019.

[93] M. Zheng, T. Li, R. Zhu, Y. Tang, M. Tang, L. Lin, and Z. Ma, "Conditional wasserstein generative adversarial network-gradient penalty-based approach to alleviating imbalanced data classification," *Information Sciences*, vol. 512, pp. 1009–1023, Feb. 2020.